REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
13-09-2010	Final Report	1-May-2005 - 30-Sep-2010

4. TITLE AND SUBTITLE
Final Report, Army Research Office Grant W911NF-05-1-0153, Multidisciplinary University Research Initiative, Training Knowledge and Skills for the Networked Battlefield

5a. CONTRACT NUMBER
W911NF-05-1-0153

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
611103

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHORS
Alice F. Healy, Lyle E. Bourne Jr., Benjamin Clegg, Bengt Fornberg, Cleotilde Gonzalez, Eric Heggestad, Robert Proctor, William D. Raymond, Carolyn J. Buck-Gengler

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
University of Colorado - Boulder
Office of Contracts and Grants
Campus Box 572, 3100 Marine Street Rm 481
Boulder, CO 80309 -0572

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
48097-MA-MUR.92

12. DISTRIBUTION AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT
The goal of our research, which has been supported by multidisciplinary university research initiative (MURI) grant W911NF-05-1-0153 from the Army Research Office, has been to construct a theoretical and empirical framework that can account for and make accurate predictions about the effectiveness of different training methods over the full range of militarily relevant tasks. The ability to predict the outcomes of different training methods on particular tasks will, as a natural by-product, point to ways to optimize training. The work performed in our project

15. SUBJECT TERMS
training knowledge and skills, retention of training, transfer of training, cognitive models of training, training principles, taxonomy of training

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
UU	UU	UU	UU		Alice Healy
					19b. TELEPHONE NUMBER: 303-492-5032

Standard Form 298 (Rev 8/98)
Prescribed by ANSI Std. Z39.18

Report Title

Final Report, Army Research Office Grant W911NF-05-1-0153, Multidisciplinary University Research Initiative, Training Knowledge and Skills for the Networked Battlefield

ABSTRACT

The goal of our research, which has been supported by multidisciplinary university research initiative (MURI) grant W911NF-05-1-0153 from the Army Research Office, has been to construct a theoretical and empirical framework that can account for and make accurate predictions about the effectiveness of different training methods over the full range of militarily relevant tasks. The ability to predict the outcomes of different training methods on particular tasks will, as a natural by-product, point to ways to optimize training. The work performed in our project falls into three interrelated categories: First, empirical studies have been conducted on (a) the development and testing of training principles, (b) the acquisition and retention of basic components of skill, and (c) levels of automation, individual differences, and team performance. Second, a taxonomic analysis of training and task types was developed and extended to include training principles and performance measures. Third, based on the first two efforts, predictive cognitive models of training effects were formulated and tested for applicability to performance by military personnel.

List of papers submitted or published that acknowledge ARO support during this reporting period. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

Bonk, W. J., & Healy, A. F. (2010). Learning and memory for sequences of pictures, words, and spatial locations: An exploration of serial position effects. American Journal of Psychology, 123, 137-168.

Bourne, L. E., Jr., Healy, A. F., Bonk, W. J., & Buck-Gengler, C. J. (in press). Intention to respond in a special way offers some protection against forgetting associations. American Journal of Psychology.

Bourne, L. E., Jr., Raymond, W. D., & Healy, A. F. (2010). Strategy selection and use during classification skill acquisition. Journal of Experimental Psychology: Learning, Memory, and Cognition, 36, 500-514.

Choi, J. M., Cho, Y. S., & Proctor, R. W. (2009). Impaired color word processing at an unattended location: Evidence from the Stroop task combined with inhibition of return. Memory & Cognition, 37, 935-944.

Durrance Blalock, L., & Clegg, B. A. (2010). Encoding and representation of simultaneous and sequential arrays in visuospatial working memory. Quarterly Journal of Experimental Psychology, 63, 856-862.

Gonzalez, C., Best, B., Healy, A. F., Kole, J. A., & Bourne, L. E., Jr. (in press). A cognitive modeling account of simultaneous learning and fatigue effects. Cognitive Systems Research.

Kole, J. A., Healy, A. F., Fierman, D. M., & Bourne, L. E., Jr. (2010). Contextual memory and skill transfer in category search. Memory & Cognition, 38, 67-82.

Lohse, K. R., Sherwood, D. E., & Healy, A. F. (2010). How changing the focus of attention affects performance, kinematics, and electromyography in dart throwing. Human Movement Science, 29, 542-555.

Miles, J. D., & Proctor, R. W. (2009). Non-intentional but not automatic: Reduction of word- and arrow-based compatibility effects by sound distractors in the same categorical domain. Experimental Brain Research, 199, 101-106.

Miles, J. D., & Proctor, R. W. (in press). Attention is required for the acquisition but not expression of new spatial associations. Journal of Experimental Psychology: Learning, Memory, and Cognition.

Miles, J. D., Yamaguchi, M., & Proctor, R. W. (2009). Dilution of the Simon effect: Categorical similarity between distractors and diluters. Attention, Perception, & Psychophysics, 71, 1598-1606.

Miles, J. D., Witt, J. K., & Proctor, R. W. (in press). Action plans produce separate Simon effects for picking up objects and transporting them. Psychological Research/Psychologische Forschung.

Overstreet, M. F., & Healy, A. F. (in press). Item and order information in semantic memory: Students' retention of the CU Fight Song lyrics. Memory & Cognition.

Proctor, R. W. (in press). Playing the Simon game: Use of the Simon task for investigating human information processing. Acta Psychologica.

Proctor, R. W., & Shao, C. (2010). Does the contribution of stimulus-hand correspondence to the auditory Simon effect increase with practice? Experimental Brain Research, 204, 131-137.

Proctor, R. W., & Vu, K.-P. L. (2009). Determinants of the benefit for consistent stimulus-response mappings in dual-task performance of three-choice tasks. Attention, Perception, & Psychophysics, 71, 1771-1781.

Proctor, R. W., & Vu, K.-P. L. (2010). Cumulative knowledge and progress in human factors. Annual Review of Psychology, 61, 623-651.

Proctor, R. W., & Vu, K.-P. L. (2010). Stimulus-response compatibility for mixed mappings and tasks with unique responses. Quarterly Journal of Experimental Psychology, 63, 320–340.

Proctor, R. W., & Vu, K.-P. L. (in press). Universal and culture-specific effects of display-control compatibility. American Journal of Psychology.

Richard, M. V., Clegg, B. A., & Seger, C. A. (2009). Implicit motor sequence learning is not represented purely in response locations. Quarterly Journal of Experimental Psychology, 62, 1516-1522.

Schneider, V. I., Healy, A. F., Barshi, I., & Kole, J. A. (in press). Following navigation instructions presented verbally or spatially: Effects on training, retention, and transfer. Applied Cognitive Psychology.

Shin, Y. K., Proctor, R. W., & Capaldi, E. J. (in press). A review of contemporary ideomotor theory. Psychological Bulletin.

Vu, K.-P. L., Ngo, T. K., Minakata, K., & Proctor, R. W. (in press). Shared spatial representations for physical locations and location words in Bilinguals' primary language. Memory & Cognition.

Wang, D. D., Proctor, R. W., & Pick, D. F. (2009). Allocation of effort as a function of payoffs for individual tasks in a multitasking environment. Behavior Research Methods, 41, 705-716.

Wohldmann, E. L., & Healy, A. F. (2010). Exploring specificity of speeded aiming movements: Examining different measures of transfer. Memory & Cognition, 38, 344-355.

Wohldmann, E. L., Healy, A. F., & Bourne, L. E., Jr. (2010). Task integration in time production. Attention, Perception, & Psychophysics, 72, 1130-1143.

Yamaguchi, M., & Proctor, R. W. (2010). Compatibility of motion information in two aircraft attitude displays for a tracking task. American Journal of Psychology, 123, 81-92.

Yamaguchi, M., & Proctor, R. W. (in press). The Simon task with multi-component responses: Two loci of response-effect compatibility. Psychological Research/Psychologische Forschung.

Young, M. D., Healy, A. F., Gonzalez, C., Dutt, V., & Bourne, L. E., Jr. (in press). Effects of training with added difficulties on RADAR detection. Applied Cognitive Psychology.

Number of Papers published in peer-reviewed journals: 29.00

(b) Papers published in non-peer-reviewed journals or in conference proceedings (N/A for none)

Bourne, L. E., Jr., & Healy, A. F. (2010). Experimental psychology. In I. B. Weiner & W. E. Craighead (Eds.), Corsini encyclopedia of psychology (4th Edition, Vol., 2, pp. 619-623). Hoboken, NJ: Wiley.

Healy, A. F., Kole, J. A., Wohldmann, E. L., Buck-Gengler, C. J., & Bourne, L. E., Jr. (in press). Data entry: A window to principles of training. In A. S. Benjamin (Ed.), Successful remembering and successful forgetting: A festschrift in honor of Robert A. Bjork. New York: Psychology Press.

Proctor, R. W., & Yamaguchi, M. (2010). Factors affecting speed and accuracy of response selection in operational environments. In D. H. Andrews, R. P. Herz, & M. B. Wolf (Eds.), Human factors issues in combat identification (pp. 31-47). Farnham, UK: Ashgate.

Yamaguchi, M., & Proctor, R. W. (in press). Perception and attention. In R. R. Hoffman, J. Szalma, M. Scerbo, P. Hancock, & R. Parasuraman (Eds.), Cambridge Handbook of Applied Perception Research.

Young, M. D., Wilson, M. L., & Healy, A. F. (2010). Improving reading skills for ESL learners using SoundSpel. In E. F. Caldwell (Ed.), Bilinguals: Cognition, education and language processing (215-227). Hauppauge, NY: Nova Science Publishers.

Number of Papers published in non peer-reviewed journals: 5.00

(c) Presentations

Anderson, L. S., Healy, A. F., Kole, J. A., & Bourne, L. E., Jr. (2010, April). The clicker technique: An effective method of teaching compression. Poster presented at the 80th Annual Convention of the Rocky Mountain Psychological Association, Denver, CO.

Anderson, L. S., Healy, A. F., Kole, J. A., & Bourne, L. E., Jr. (2010, November). The clicker technique: An effective way to compress teaching time. Paper to be presented at the 51st Annual Meeting of the Psychonomic Society, St. Louis, MO.

Baroni, G., Yamaguchi, M., & Proctor, R. W. (2010, November). Transfer from a color-mapping task to a Simon task for shapes. Poster to be presented at the 51st Annual Meeting of the Psychonomic Society, St. Louis, MO.

Bourne, L. E., Jr., Healy, A. F., Bonk, W. J., & Buck-Gengler, C. J. (2009, November). Prospective memory offers some protection against forgetting associated items. Paper presented at the 50th Annual Meeting of the Psychonomic Society, Boston, MA.

Buck-Gengler, C. J., Raymond, W. D., Healy, A. F., & Bourne, L. E., Jr. (2010, March). Modeling a visual search task with a secondary task in IMPRINT. Poster presented at the 19th Annual Conference on Behavior Representation in Modeling and Simulation (BRIMS), Charleston, SC.

Clegg, B. A. & Heggestad, E. D. (2010, April). Experiments on levels of automation, individual differences, and team performance. Paper presented at the 2010 Ellis-Battig Memory Symposium: Optimizing the training of knowledge and skills: A review of accomplishments from the Multidisciplinary University Research Initiative (MURI) on training, 80th Annual Convention of the Rocky Mountain Psychological Association, Denver, CO.

Durrance Blalock, L., & Clegg, B. A. (2009, November). Impact of item familiarity on short-term consolidation in visual working memory. Poster presented at the 50th Annual Meeting of the Psychonomic Society, Boston, MA.

Durrance Blalock, L., & Clegg, B. A. (2010, April). The impact of load on simulated driving and situational awareness. Poster presented at the 80th Annual Meeting of the Rocky Mountain Psychological Association, Denver, CO.

Gonzalez, C. (2010, April). Modeling training performance in ACT-R. Paper presented at the 2010 Ellis-Battig Memory Symposium: Optimizing the training of knowledge and skills: A review of accomplishments from the Multidisciplinary University Research Initiative (MURI) on training, 80th Annual Convention of the Rocky Mountain Psychological Association, Denver, CO.

Healy, A. F. (2009, October). Principles of training. Invited paper presented at the Workshop to Explore Issues and Mitigation Strategies for Long Term Retention of Military Expertise. Mesa, Arizona.

Healy, A. F. (2010, April). Experiments on development and testing of training principles. Paper presented at the 2010 Ellis-Battig Memory Symposium: Optimizing the training of knowledge and skills: A review of accomplishments from the Multidisciplinary University Research Initiative (MURI) on training, 80th Annual Convention of the Rocky Mountain Psychological Association, Denver, CO.

Healy, A. F., & Cunningham, T. F. (2009, November). Detection of letter and letter sequence targets while processing prose. Poster presented at the 50th Annual Meeting of the Psychonomic Society, Boston, MA.

Jones, M., Worthy, D. A., Ketels, S. L., & Otto, A. R. (2010, August). The phenomenology of multiple learning systems. Paper presented at the Ninth Annual Summer Interdisciplinary Conference, Bend, OR.

Ketels, S. L., Healy, A. F., Wickens, C. D., Buck-Gengler, C. J., & Bourne, L. E., Jr. (2010, April). Spatial list learning and decision making in the fusion paradigm. Poster presented at the 80th Annual Convention of the Rocky Mountain Psychological Association, Denver, CO.

Ketels, S., & Jones, M. (2010, August). Language is not always helpful: Labels do not facilitate the learning of information-integration category structures. Poster presented at the Ninth Annual Summer Interdisciplinary Conference. Bend, OR.

Kole, J. A., & Healy, A. F. (2009, November). Long-term retention of knowledge about friends, family, and unfamiliar individuals. Poster presented at the 50th Annual Meeting of the Psychonomic Society, Boston, MA.

Kole, J. A., & Healy, A. F. (2010, April). Memory for facts about people: Familiarity, relatedness, degree of genetic similarity, and gender congruency. Poster presented at the 80th Annual Convention of the Rocky Mountain Psychological Association, Denver, CO.

Kole, J. A., & Healy, A. F. (2010, November). Memory for details about people: Familiarity, relatedness, and gender congruency. Poster to be presented at the 51st Annual Meeting of the Psychonomic Society, St. Louis, MO.

Lohse, K. R., Healy, A. F., & Sherwood, D. E. (2009, November). Task-level and effector-level representations in intermanual transfer of motor skills. Poster presented at the 50th Annual Meeting of the Psychonomic Society, Boston, MA.

Lohse, K. R., Sherwood, D. E., & Healy, A. F. (2010, April). How changing the focus of attention affects performance, kinematics, and electromyography in dart throwing. Paper presented at the 80th Annual Convention of the Rocky Mountain Psychological Association, Denver, CO.

Lohse, K. R., Sherwood, D. E., & Healy, A. F. (2010, June). How changing the focus of attention affects performance, kinematics, and electromyography. Paper presented at NASPSA (North American Society for Psychology of Sport and Physical Activity) Conference 2010, Tucson, AZ.

Lohse, K. R., Sherwood, D. E., & Healy, A. F. (2010, June). The influence of attention on learning and performance: Two experiments in isometric force production. Poster presented at NASPSA (North American Society for Psychology of Sport and Physical Activity) Conference 2010, Tucson, AZ.

McCormick, B., & Healy, A. F. (2010, April). Words and symbols use different working memory resources in a navigational task. Poster presented at the 80th Annual Convention of the Rocky Mountain Psychological Association, Denver, CO.

Miles, J. (2009, September) Conceptual mediation of spatial stimulus-response compatibility effects. Invited Presentation at The European Society for Cognitive Psychology Conference, Krakow, Poland.

Miles, J. D., & Proctor, R. W. (2009, November). Object and goal locations contribute additively to the Simon effect. Poster presented at the 50th annual meeting of the Psychonomic Society, Boston, MA.

Miles, J. D., & Proctor, R. W. (2010, April). Attention is required for the acquisition of new spatial associations but not their expression. Paper presented at the Rocky Mountain Psychological Association Conference, Denver, CO.

Miles, J. D., & Proctor, R. W. (2010, November). Feature attraction between controlled objects and targets. Poster to be presented at the 51st Annual Meeting of the Psychonomic Society, St. Louis, MO.

Mong, H. M., Clegg, B. A., & Seger, C. A. (2009, November). Implicit learning in enumeration with a process dissociation knowledge test. Poster presented at the 50th Annual Meeting of the Psychonomic Society, Boston, MA.

Mong, H. M., Murchison, N., & Clegg, B. A. (2010, April). Effects of spatial training on hazard detection with simulated head up displays. Poster presented at the 80th Annual Meeting of the Rocky Mountain Psychological Association, Denver, CO.

Overstreet, M. F., & Healy, A. F. (2010, April). Item and order information in semantic memory: Students' retention of the CU fight song. Poster presented at the 80th Annual Convention of the Rocky Mountain Psychological Association, Denver, CO.

Proctor, R. W. (2010, April). Experiments on acquisition and retention of basic components of skill. Paper presented in the 2010 Ellis-Battig Memory Symposium, Optimizing the Training of Knowledge and Skills: A Review of Accomplishments from the Multidisciplinary University Research Initiative (MURI) on Training. Rocky Mountain Psychological Association Conference, Denver, CO.

Proctor, R. W., & Yamaguchi, M. (2009, November). A vector model of the Simon effect. Paper presented at the 50th annual meeting of the Psychonomic Society, Boston, MA.

Proctor, R. W., Yamaguchi, M., Gonzalez, C., & Dutt, V. (2010, August). Spatial compatibility effects in a complex task environment. Paper to be presented at the 118th Annual Convention of the American Psychological Association. San Diego, CA.

Raymond, W. D., & Buck-Gengler, C. J. (2010, April). Modeling training performance in IMPRINT. Paper presented at the 2010 Ellis-Battig Memory Symposium: Optimizing the training of knowledge and skills: A review of accomplishments from the

Multidisciplinary University Research Initiative (MURI) on training, 80th Annual Convention of the Rocky Mountain Psychological Association, Denver, CO.

Rotterman, A., Vu, K.-P. L., & Proctor, R. W. (2009, November). Effects of practice with noncorresponding location words on the Simon task. Poster presented at the 50th annual meeting of the Psychonomic Society, Boston, MA.

Schneider, V. I., Healy, A. F., & Barshi, I. (2010, November). Learning specificity: modality transfer in following navigation instructions. Paper to be presented at the 51st Annual Meeting of the Psychonomic Society, St. Louis, MO.

Schneider, V. I., Healy, A. F., Barshi, I., McCormick, B., & Bourne, L. E., Jr. (2009, November). Effects of presentation order during training to follow navigation instructions. Poster presented at the 50th Annual Meeting of the Psychonomic Society, Boston, MA.

Wohldmann, E. L., & Healy, A. F. (2010, November). Indices of transfer: Learning can transfer but still be specific. Poster to be presented at the 51st Annual Meeting of the Psychonomic Society, St. Louis, MO.

Yamaguchi, M., & Proctor, R. W. (2009, November). Validating the vector model assumptions: Extension to a Simon task with multidimensional responses. Poster presented at the 50th annual meeting of the Psychonomic Society, Boston, MA.

Yamaguchi, M., & Proctor, R. W. (2010, April). A vector representation underlying response-selection processes. Poster presented at the Rocky Mountain Psychological Association Conference, Denver, CO.

Yamaguchi, M., & Proctor, R. W. (2010, November). Automatic implementation of task-defined rules: Active maintenance or memory retrieval? Poster to be presented at the 51st Annual Meeting of the Psychonomic Society, St. Louis, MO.

Young, M. D., Healy, A. F., & Bourne, L. E., Jr. (2009, November). Training and transfer of an artificial grammar. Poster presented at the 50th Annual Meeting of the Psychonomic Society, Boston, MA, November 20, 2009.

Young, M. D., Healy, A. F., & Bourne, L. E., Jr. (2010, April). Artificial grammar learning: Retention and transfer. Poster presented at the 80th Annual Convention of the Rocky Mountain Psychological Association, Denver, CO.

Zhang, Y., & Proctor, R. W. (2009, November). Comparison of affective and spatial Simon effects for word stimuli. Poster presented at the 50th annual meeting of the Psychonomic Society, Boston, MA.

Number of Presentations: 44.00

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts): 0

Peer-Reviewed Conference Proceeding publications (other than abstracts):

Buck-Gengler, C. J., Raymond, W. D., Healy, A. F., & Bourne, L. E., Jr. (2010). Modeling a visual search task with a secondary task in IMPRINT. In Proceedings of the Nineteenth Annual Conference on Behavior Representation in Modeling and Simulation, Charleston, SC, March 22-25, 2010. Available on CD-ROM.

Clegg, B. A., Heggestad, E. D., & Durrance Blalock, L. D. (in press). The influences of automation and trainee aptitude on training effectiveness. In Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting, San Francisco, CA, September 27-October 1, 2010. Human Factors and Ergonomics Society.

Dunston, P. S., Proctor, R. W., Su, X., Yamaguchi, M., Wang, X., & Chen, R. (2010). Principles for utilization of construction equipment operator training simulators. In J. Ruwanpura, Y. Mohamed, & S. Lee (Eds.), Innovation for Reshaping Construction Practice, Proceedings of the 2010 Construction Research Congress, American Society of Civil Engineers, Volume 2, May 8-10, Banff, Alberta, Canada, 1039-1046.

Durrance Blalock, L., Sawyer, B. D., Kiken, A. & Clegg, B. A. (2009). The impact of load on dynamic versus static situational knowledge while driving. Proceedings of 53rd Annual Meeting of the Human Factors and Ergonomics Society, San Antonio, TX.

Dutt, V., Yamaguchi, M., Gonzalez, C., & Proctor, R. W. (2009). An instance-based learning model of stimulus-response compatibility effects in mixed location-relevant and location-irrelevant tasks. In A. Howes, D. Peebles, & R. Cooper (Eds.), 9th International Conference on Cognitive Modeling – ICCM2009. July 24-26, 2009. Manchester, UK: University of Manchester.

Gonzalez, C., & Dutt, V. (in press). Instance-based learning models of training. In Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting, San Francisco, CA, September 27-October 1, 2010. Human Factors and Ergonomics Society.

Gonzalez, C., Dutt, V., Healy, A. F., Young, M. D., & Bourne, L. E., Jr. (2009). Comparison of instance and strategy models in ACT-R. In A. Howes, D. Peebles, & R. Cooper (Eds.), 9th International Conference on Cognitive Modeling – ICCM2009. Manchester, UK.

Wickens, C. D., Ketels, S. L., Healy, A. F., Buck-Gengler, C. J., & Bourne, L. E., Jr. (in press). The anchoring heuristic in intelligence integration: A bias in need of de-biasing. In Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting, San Francisco, CA, September 27-October 1, 2010. Human Factors and Ergonomics Society.

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts): 8

(d) Manuscripts

Barshi, I., & Healy, A. F. (2010). The effects of spatial representation on memory for verbal navigation instructions. Manuscript submitted for publication.

Clegg, B. A., & Heggestad, E. D. (2010). The influences of automation and trainee aptitude on training effectiveness. Manuscript submitted for publication.

Dutt, V., Gonzalez, C., Yamaguchi, M., & Proctor, R. W. (2010). Instance-based learning models of SRC and Simon effects. Manuscript submitted for publication.

Healy, A. F., & Bourne, L. E., Jr. (2010). Principles of training. Manuscript submitted for publication.

Healy, A. F., & Cunningham, T. (2010). Detecting letters and words in prose passages. Manuscript submitted for publication.

Healy, A. F., Wohldmann, E. L., & Bourne, L. E., Jr. (2010). Does practice with a defective mouse influence subsequent speeded aiming performance? A test of global inhibition. Manuscript submitted for publication.

Koch, I., Schuch, S., Vu, K.-P. L., & Proctor, R. W. (2010). Response discriminability in task switching—Dissociating effects of anatomical and spatial response separation. Manuscript submitted for publication.

Kole, J. A., & Healy, A. F. (2010). Memory for details about people: Familiarity, relatedness, and gender congruency. Manuscript submitted for publication.

Krech Thomas, H., & Healy, A. F. (2010). A comparison of rereading benefits in first- and second-language reading. Manuscript submitted for publication.

Lohse, K. R., & Healy, A. F. (2009). Exploring the contributions of declarative and procedural information to training: A test of the procedural reinstatement principle. Manuscript submitted for publication.

Lohse, K. R., Healy, A. F., & Sherwood, D. E. (2010). Mental practice in the intermanual transfer of motor skills. Manuscript submitted for publication.

Lohse, K. R., Sherwood, D. E., & Healy, A. F. (2010). Neuromuscular effects of shifting the focus of attention in a simple force production task. Manuscript submitted for publication.

Raymond, W. D., Healy, A. F., & Bourne, L. E., Jr. (2010). The MURI training taxonomy. Manuscript submitted for publication.

Raymond, W. D., Healy, A. F., & McDonnel, S. J. (2010). Pairing words with syntactic frames: Syntax, semantics, and count-mass usage. Manuscript submitted for publication.

Yamaguchi, M., & Proctor, R. W. (2010). Action-based selection across feature dimensions. Manuscript submitted for publication.

Yamaguchi, M., & Proctor, R. W. (2010). Automaticity without training: The role of memory retrieval in automatic implementation of task-defined rules. Manuscript submitted for publication.

Yamaguchi, M., & Proctor, R. W. (2010). Multidimensional vector model of stimulus-response compatibility. Manuscript submitted for publication.

Yamaguchi, M., & Proctor, R. W. (2010). Response competition across feature dimensions. Manuscript submitted for publication.

Number of Manuscripts: 18.00

Patents Submitted

Patents Awarded

Graduate Students

NAME	PERCENT_SUPPORTED
Michael Young	0.32
Shaw Ketels	0.54
Keith Lohse	0.21
Lindsay Anderson	0.21
Blu McCormick	0.18
Motonori Yamaguchi	0.21
Lisa Durance Blalock	0.14
Varun Dutt	0.80
Heather Mong	0.14
FTE Equivalent:	**2.75**
Total Number:	**9**

Names of Post Doctorates

NAME	PERCENT_SUPPORTED
Carolyn Buck-Gengler	0.50
William Raymond	0.82
James Kole	0.76
James Miles	0.50
FTE Equivalent:	**2.58**
Total Number:	**4**

Names of Faculty Supported

NAME	PERCENT_SUPPORTED	National Academy Member
Alice Healy	0.38	No
Lyle Bourne	0.38	No
Bengt Fornberg	0.07	No
Matt Jones	0.07	No
Robert Proctor	0.30	No
Benjamin Clegg	0.14	No
Eric Heggestad	0.14	No
Cleotilde Gonzalez	0.20	No
FTE Equivalent:	**1.68**	
Total Number:	**8**	

Names of Under Graduate students supported

NAME	PERCENT_SUPPORTED
Michael Overstreet	0.02
Amir Ahmed	0.12
Kalyn McGinnis	0.11
Hannah St. Louis	0.02
FTE Equivalent:	**0.27**
Total Number:	**4**

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period:	1.00
The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:......	1.00
The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:......	0.00
Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):	0.00
Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:	0.00
The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense	0.00
The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields:	0.00

Names of Personnel receiving masters degrees

NAME	
Total Number:	

Names of personnel receiving PHDs

NAME	
Michael Young	
Motonori Yamaguchi	
Lisa Durrance Blalock	
Total Number:	**3**

Names of other research staff

NAME	PERCENT_SUPPORTED	
Michael Overstreet	0.20	No
FTE Equivalent:	**0.20**	
Total Number:	**1**	

Sub Contractors (DD882)

1 a. Purdue University

1 b. Sponsored Programs Services
Young Hall, 302 Wood Street
West Lafayette IN 479072108

Sub Contractor Numbers (c):
Patent Clause Number (d-1):
Patent Date (d-2):
Work Description (e):
Sub Contract Award Date (f-1): 5/1/2005 12:00:00AM
Sub Contract Est Completion Date(f-2): 9/30/2010 12:00:00AM

---

1 a. Purdue University

1 b. Sponsored Program Services
610 Purdue Mall
West Lafayette IN 479072040

Sub Contractor Numbers (c):
Patent Clause Number (d-1):
Patent Date (d-2):
Work Description (e):
Sub Contract Award Date (f-1): 5/1/2005 12:00:00AM
Sub Contract Est Completion Date(f-2): 9/30/2010 12:00:00AM

---

1 a. Carnegie Mellon University

1 b. Office of Sponsored Programs
Carnegie Mellon University
Pittsburgh PA 15213

Sub Contractor Numbers (c):
Patent Clause Number (d-1):
Patent Date (d-2):
Work Description (e):
Sub Contract Award Date (f-1): 5/1/2005 12:00:00AM
Sub Contract Est Completion Date(f-2): 9/30/2010 12:00:00AM

---

1 a. Carnegie Mellon University

1 b. Office of Sponsored Programs
5000 Forbes Avenue
Pittsburgh PA 152133890

Sub Contractor Numbers (c):
Patent Clause Number (d-1):
Patent Date (d-2):
Work Description (e):
Sub Contract Award Date (f-1): 5/1/2005 12:00:00AM
Sub Contract Est Completion Date(f-2): 9/30/2010 12:00:00AM

---

1 a. Purdue University

1 b. 155 S. Grant Street
Sponsored Program Services
West Lafayette IN 47907-2108

Sub Contractor Numbers (c): 1541622
Patent Clause Number (d-1):
Patent Date (d-2):
Work Description (e): Conduct experiments on the acquisition and retention of basic components of skill
Sub Contract Award Date (f-1): 5/1/2005 12:00:00AM
Sub Contract Est Completion Date(f-2): 9/30/2010 12:00:00AM

---

1 a. Carnegie Mellon University

1 b. 401 Warner Hall
5000 Forbes Avenue
Pittsburgh PA 15213-3890

Sub Contractor Numbers (c): 1541620
Patent Clause Number (d-1):
Patent Date (d-2):
Work Description (e): Develop ACT-R models of training
Sub Contract Award Date (f-1): 5/1/2005 12:00:00AM
Sub Contract Est Completion Date(f-2): 9/30/2010 12:00:00AM

---

1 a. Colorado State University - Ft. Collins

1 b. Sponsored Programs
601 South Howes St., 408 University Se
Fort Collins CO 805232002

Sub Contractor Numbers (c): 1541621
Patent Clause Number (d-1):
Patent Date (d-2):
Work Description (e): Conduct experiments on levels of automation, individual differences, and team performance.
Sub Contract Award Date (f-1): 5/1/2005 12:00:00AM
Sub Contract Est Completion Date(f-2): 9/30/2010 12:00:00AM

---

1 a. Colorado State University - Ft. Collins

1 b. Office of Sponsored Programs
Colorado State University
Fort Collins CO 805232002

Sub Contractor Numbers (c): 1541621
Patent Clause Number (d-1):
Patent Date (d-2):
Work Description (e): Conduct experiments on levels of automation, individual differences, and team performance.
Sub Contract Award Date (f-1): 5/1/2005 12:00:00AM
Sub Contract Est Completion Date(f-2): 9/30/2010 12:00:00AM

Inventions (DD882)

FINAL REPORT

Army Research Office Grant W911NF-05-1-0153
Multidisciplinary University Research Initiative
Training Knowledge and Skills for the Networked Battlefield

Alice F. Healy and Lyle E. Bourne, Jr.
Principal Investigators

Benjamin A. Clegg, Bengt Fornberg, Cleotilde Gonzalez, Eric D. Heggestad,
and Robert W. Proctor
Co-Investigators

Carolyn J. Buck-Gengler and William D. Raymond
Research Associates

Foreword

The goal of our research, which has been supported by multidisciplinary university research initiative (MURI) grant W911NF-05-1-0153 from the Army Research Office, has been to construct a theoretical and empirical framework that can account for and make accurate predictions about the effectiveness of different training methods over the full range of militarily relevant tasks. The ability to predict the outcomes of different training methods on particular tasks will, as a natural by-product, point to ways to optimize training. The work performed in our project falls into three interrelated categories: First, empirical studies have been conducted on (a) the development and testing of training principles, (b) the acquisition and retention of basic components of skill, and (c) levels of automation, individual differences, and team performance. Second, a taxonomic analysis of training and task types was developed and extended to include training principles and performance measures. Third, based on the first two efforts, predictive cognitive models of training effects were formulated and tested for applicability to performance by military personnel.

Table of Contents

	Page
A. Empirical Studies of Training	3
1. Development and Testing of Training Principles	3
a. Problem studied	3
b. Important results	3
2. Acquisition and Retention of Basic Components of Skill	5
a. Problem studied	5
b. Important results	5
3. Levels of Automation, Individual Differences, and Team Performance	6
a. Problem studied	6
b. Important results	7
B. Taxonomy	7
1. Problem studied	7
2. Important results	8
C. Cognitive Models of Training	8
1. ACT-R Models	8
a. Problem studied	9
b. Important results	9
2. IMPRINT Models	10
a. Problem studied	10
b. Important results	11
3. Model Comparison and Evaluation	12
a. Problem studied	12
b. Important results	12
D. Summary	13
E. Bibliography	14

A. Empirical Studies of Training

1. Development and Testing of Training Principles

a. Problem studied. Research was aimed at identifying and empirically supporting training principles for procedural and declarative memory skills. These principles can provide guidelines to trainers that will enhance the effectiveness of the training they perform. The principles investigated in experiments on the development and testing of training principles were conducted on a range of issues, including (a) generality across tasks of individual principles, (b) tests of multiple principles in a single task, (c) tests of principles in complex, dynamic environments, and (d) development and testing of new principles.

b. Important results. The following 20 principles (in abbreviated form) have been among those investigated in this experimental research using a variety of tasks and paradigms:

Bilateral Transfer: For spatial motor skills, there is more transfer from the dominant to the non-dominant hand than in the opposite direction (Lohse, Healy, & Sherwood, 2009).

Cognitive Antidote: Adding cognitive complications to a routine task overcomes the decline in accuracy due to fatigue (Kole, Healy, & Bourne, 2008).

Depth of Processing: Activities during training that promote deep and elaborate processing enhance durability of training (Healy, Kole, Wohldmann, Buck-Gengler, & Bourne, in press).

Dual Coding: Retention is best for items encoded both verbally and spatially (Bonk & Healy, 2010).

List Length: Retention of a given item in a list is better for short lists than for long lists (Bonk & Healy, 2010).

Memory Constriction: The time span from which memories can be retrieved shrinks as stress increases (Staal, Bolton, Yaroush, & Bourne, 2008).

Mental Practice: Mental practice promotes task-level representations but not effector-level representations of motor skill (Lohse et al., 2009).

Mental Rehearsal: Mental rehearsal can retard forgetting and promote transfer of training to a larger extent than can physical rehearsal, which suffers from motoric interference (Wohldmann, Healy, & Bourne, 2007, 2008a).

Optimal Modality Use: Learning is better when information is seen than when it is read, and it is best when the information is both read and seen (Schneider, Healy, Buck-Gengler, Barshi, & Bourne, 2008).

Positive Focusing: Regularities obeying complex rules can sometimes be best appreciated with only positive exemplars, rather than both positive and negative exemplars (Young, 2010).

Procedural Reinstatement: Specificity (limited transfer) occurs for tasks based primarily on procedural information, or skill, whereas generality (robust transfer) occurs for tasks based primarily on declarative information, or facts. Alternatively, duplicating procedures required during learning facilitates later retention and transfer (Healy, Wohldmann, Kole, Schneider, Shea, & Bourne, in press; Kole, Healy, Fierman, & Bourne, 2010).

Retrieval Distraction: Retention is best when tested with minimal distraction (Bonk & Healy, 2010).

Retrieval Induced Forgetting: Retrieval of information from memory can cause forgetting of related information not retrieved (Kole & Healy, 2008).

Serial Position: Retention is best for items at the start of a list (primacy advantage) and at the end of a list (recency advantage) (Bonk & Healy, 2010; Ketels, Healy, Wickens, Buck-Gengler, & Bourne, 2010; Wickens, Ketels, Healy, Buck-Gengler, & Bourne, in press).

Specificity of Training: Retention and transfer are depressed when conditions of learning differ from those during subsequent testing (Wohldmann & Healy, 2010; Wohldmann, Healy, & Bourne, 2010).

Strategic Use of Knowledge: Learning and memory are facilitated whenever pre-existing knowledge can be employed, possibly as a mediator, in the process of acquisition (Kole & Healy, 2007, 2010).

Testing: A test can strengthen a person’s knowledge of material as much as, or possibly even more than, can further study (Anderson, Healy, Kole, & Bourne, 2010).

Training Compression: Training can be truncated by eliminating practice on known facts (Anderson et al., 2010).

Training Difficulty: Any condition that causes difficulty during learning may facilitate later retention and transfer (Young, Healy, Gonzalez, Dutt, & Bourne, in press).

Variability of Practice: Variable practice conditions typically yield larger transfer effects compared with constant practice conditions (Wohldmann & Healy, 2010; Wohldmann, Healy, & Bourne, 2008b).

The literature documenting the full set of training principles is summarized in a technical report (Healy, Schneider, & Bourne, 2010). A quantitative version of selected training principles is provided in a second technical report (Bourne, Raymond, & Healy, 2010). The publications and presentations resulting from the MURI research providing the empirical validation of the training principles are summarized in a third technical report (Healy & Bourne, 2010).

2. Acquisition and Retention of Basic Components of Skill

a. Problem studied. The goal of this part of the Training MURI was to isolate the perceptual, cognitive, and motor components of skills and examine factors that affect acquisition and transfer of these skills. Much of this work focused on response selection, the processes involved in deciding which responses to make to stimuli in particular situations. Examining skill acquisition in tasks that stress response selection is important because it is the aspect of skill that benefits the most from training and practice (Welford, 1976). Our work focused on three domains of basic skills: (a) transfer of newly acquired associations, (b) training with mixed mappings and tasks, and (c) performance in settings that require multitasking.

b. Important results. Perhaps the most striking outcome of our transfer studies is how easy it is to overcome or counteract effects of pre-existing performance biases. The benefit for spatial correspondence is eliminated by less than 100 trials of practice with an incompatible spatial mapping. With larger amounts of practice, the transfer task shows reversal of the Simon effect (faster responding for stimuli and responses compatible in location when location is irrelevant to the task at hand) to favor the practiced incompatible stimulus-response (S-R) relation. Nevertheless, another important aspect of transfer of learning is its limitations. In our transfer studies, we found that the transfer effect is larger when the practice and transfer contexts are similar than when they are not, with respect to stimulus modalities (visual or auditory), the types of stimulus mode, the response mode, and spatial dimensions of stimuli and responses. The results are consistent with the *specificity of training principle*. At the same time, the results of our transfer studies are largely consistent with the MURI framework (performance shaping function), in which amount of transfer is determined by number of practice trials, learning rate, contextual similarity in training and transfer contexts, and time passage.

For the mixed mappings/tasks domain, the sets (or readiness) to perform each component task are active concurrently, unlike in the practice/transfer domain. When two or more tasks or mappings of stimuli to responses are mixed, such that performers are uncertain about which one will be in effect on a particular trial, responding is slower overall and the stimulus-response mappings in effect for one task intrude on performance of the other task. We found considerable evidence for context similarity in this case as well: When each task had distinct responses, the costs associated with mixing were much less than when the tasks shared responses. Performance suffers even more when there is uncertainty about both which task to perform (should I respond to color or location?) and which mapping to use (if I am to respond to location, do I respond compatibly or incompatibly?), and sequential effects as a function of whether the task/mapping switches

from that of the preceding trial are prominent. For this situation, we found that the practice and sequential effects could be fit well by an ACT-R model based on the idea that responses come to be based on retrieval of previous instances (Gonzalez, Lerch, & Lebiere, 2003).

Often, not only does one have to be prepared to perform one of two or more tasks, but multitasking demands require that the tasks be performed concurrently. Using a dual-task environment to investigate practice and transfer of a primary task, we obtained additional evidence consistent with the instance-based learning theory: Attention was required for acquisition of new spatial associations of stimuli and responses, but not for transfer of this learning, implying that the transfer effect reflects “automatic retrieval” of the learned skills. In other dual-task studies we found that even with very highly compatible individual tasks, practice is not sufficient to overcome interference associated with having to perform the two tasks in close temporal proximity. Using practice and transfer in a synthetic work environment involving four distinct tasks, we found that participants were sensitive to changes in payoffs in allocating their efforts among the tasks but continued to show residual effects of the prior payoff schedule.

Our research has shown that there are benefits of applying individual principles in the training of specific tasks. However, this training is not isolated and can suffer from interference from components within a task or between tasks. We have identified specific factors that influence the learning and transfer of S-R associations and how they are impacted by task switching and multitasking.

The details of the experimental research in the MURI on acquisition and retention of basic components of skill are summarized in a technical report (see Proctor, Yamaguchi, & Miles, 2010).

3. Levels of Automation, Individual Differences, and Team Performance

a. Problem studied. Although the scientific knowledge on what automation is and how it can aid human operators is flourishing, there is still much to learn about the role of automation in training. We found no previous research that has looked at how operator individual differences might relate successful incorporation of automation into training programs.

Given the importance of interactions between individual differences in cognitive ability and forms of training on training outcomes, one imperative question within our research was to examine how aptitude and training type interact, with training type defined by levels of automation.

Specifically, the key goals and objectives guiding our research were: (a) to examine the role of automation in skill learning and (b) to determine whether the aptitude of the learner interacts with presence of automation to influence the effectiveness of training.

b. Important results. Across a series of experiments, our data consistently show no benefits for learning from the presence of automation during training and frequent situations in which the presence of automation can impair learning. Although automation does assist novice operators early in training, it apparently often does so at a cost to the degree of learning that occurs during training. When automation support was removed, costs were seen for those trained with automatically-initiated automation.

An aptitude-automation interaction was observed, such that automation reduced the relationship between trainee intelligence and training performance. This presence of an aptitude-automation interaction, shown in this project for the first time, suggests the effects of automation on training are greater for lower aptitude individuals. Although supplying greater support to such users, the presence of automation may be masking differences between individuals and at the same time impairing their ability to acquire fundamental knowledge about the operation of the system. These are clearly matters of potential practical importance within numerous training situations. Our results suggest that the effectiveness of automation in training will vary not only by the type of automation and the task, but also by the aptitude of the operator.

Within the context of the current MURI effort is the *specificity of training principle*, in which learning and transfer are reduced when conditions within training differ from those encountered within a test (e.g., Healy & Bourne, 1995; Healy et al., 1993). Such a principle would suggest that changes to the nature of the task induced through the use of automation provide sufficient differences to impact performance when automation is withdrawn.

One simple solution to reducing reliance on automation, gradually withdrawing such support, proved ineffective in our research. Although there may be other approaches that serve to inoculate individuals against over-reliance on automation, these remain topics for future research.

The details of the experimental research in the MURI on automation and effective training are summarized in a technical report (see Clegg & Heggestad, 2010).

B. Taxonomy

1. Problem Studied

The goal of the Training MURI was to quantify the effects on performance of different training methods for complex military tasks. The extensive range of variables that can affect training efficacy and the multiplicity of tasks that may require training prevent an exhaustive quantification of training outcomes for specific tasks and training scenarios. In order to render the study of training effects tractable and to guide research, we developed a multi-dimensional taxonomy, which provides a framework by which training effects can be assessed and predicted for any task. The taxonomy we have developed involves a four-dimensional decomposition of the training space. It includes separate dimensions of classification for task description, training procedure, and the

context and assessment of task performance. The training principles are considered the fourth dimension. The first three dimensions are structured as hierarchical feature decompositions.

2. Important Results

The task decomposition adopted for the MURI builds on taxonomies like the Roth (1992) taxonomy of abilities, introducing a finer classification of abilities, while keeping the number of taxa tractable. Taxa were selected to capture the cognitive processing of stimuli, which was considered to be central, both because of the military's primary desire to optimize training for the networked battlefield and because most empirical studies conducted for the MURI have largely been designed to explore cognitive processing, with concomitant perceptual and psychomotor processes. In information processing tasks inputs are initially processed using perceptual and attentional abilities. Information is further synthesized with higher-order cognitive processes and memory, and output responding is planned. Finally, a psychomotor response in produced. This sequential processing cycle is reflected in the taxonomy.

The training dimension covers variables that capture the method of instruction and the types of activities performed during learning. The two major pieces in the decomposition of task learning in the MURI taxonomy are pedagogy and practice. Pedagogy captures the method of task instruction. The practice taxa are used to describe the nature of practice performed during training. Although the set of parameter values selected for inclusion in the MURI taxonomy are intended to allow an analysis of most training scenarios, additional pedagogy and practice parameters may be added to the taxonomy when they become necessary.

The performance dimension of the MURI taxonomy incorporates the two components of performance context and performance assessment. Performance context covers the conditions of and delay to post-training performance, relative to training. Performance assessment specifies measures of performance. The Kraiger, Ford, and Salas (1993) classification of learning outcomes forms the basis for the MURI performance assessment taxonomy, which includes separate taxa for assessing the acquisition of knowledge and skills, as well as attitudinal changes. Having quantified the outcome of a particular training scenario, the effectiveness of training can be measured by comparing post-training performance with performance before or at the beginning of training, using an accepted measure of training, such as the training effectiveness ratio (Wickens & Holland, 2000). Performance results can then feed back to further training design.

The details of the MURI training taxonomy are summarized in a technical report (see Raymond, Healy, & Bourne, 2010).

C. Cognitive Models of Training

1. ACT-R Models

a. *Problem studied*. We studied the cognitive functions involved in different training principles and in a variety of tasks. We relied on the ACT-R cognitive architecture (Anderson & Lebiere, 1998). The main models developed include:

1) Models of fatigue effects in a data entry task;
2) Models of stimulus-response compatibility (SRC) and Simon effects;
3) Models of dynamic visual detection in the RADAR task.

Review of these models demonstrates the benefits of using computational modeling to develop an understanding of the learning process in a variety of tasks, how they are linked to various training principles, and the utility of the models in predicting learning effects.

b. Important results. We relied on the ACT-R cognitive architecture (Anderson & Lebiere, 1998) to develop computational models in three different projects and tasks. The first project investigated fatigue effects in a data entry task (Gonzalez, Best, Healy, Bourne, & Kole, 2010). The empirical studies examined training principles such as *specificity of training, procedural reinstatement,* and *depth of processing* (Kole et al., 2008). The data entry task required subjects to see a four-digit number and then type it on the computer. Experiments involved long sessions with several blocks of many of these numbers. Typing accuracy and speed were the main measures of performance. The ACT-R cognitive model developed for the data entry task proposed a theory of fatigue that explained the effects found in several empirical data sets: Both affective and cognitive processes decay with extended time spent on the task, producing faster performance but increased errors in the task (Fu, Gonzalez, Healy, Kole, & Bourne, 2006; Gonzalez et al., 2010; Gonzalez, Fu, Healy, Kole, & Bourne, 2006;).

A major conclusion from the work in the MURI was the robustness of the Instance-Based Learning Theory (IBLT; Gonzalez et al., 2003). The IBLT, which relies on some ACT-R mechanisms, provides an approach to modeling learning based on experience and exploration. The IBLT characterizes learning as storing a sequence of action-outcome links produced by experienced events through a feedback-loop process of human and environment interactions in memory. This process increases knowledge and allows decisions to improve as experience accumulates in memory. A demonstration of the development of IBLT models of training involved the SRC and Simon effects (Dutt, Gonzalez, Yamaguchi, & Proctor, 2010; Yamaguchi, Dutt, Gonzalez, & Proctor, 2010). The SRC effect is the faster response when both stimulus and response locations correspond than when they do not. The effect is so robust that it is found even when stimulus location is irrelevant to the task, a variation known as the *Simon Effect* (Simon, 1990). Thus, a distinction between the SRC and Simon effects is made on the basis of whether the stimulus locations are relevant or irrelevant to the task. Both the SRC and Simon effects occur for visual and tactile stimuli, verbal and nonverbal symbols that convey location information (e.g., location words; Proctor, Yamaguchi, Zhang, & Vu, 2009), a variety of response modes (e.g., a steering wheel), and in more complex task environments such as flight operations (Yamaguchi & Proctor, 2006). We provided an explanation of the observed SRC/Simon effects using an IBLT model (Dutt et al., 2010; Dutt, Yamaguchi, Gonzalez, & Proctor, 2009). The model predicts learning and performance from experiments where human participants performed mixed Simon and

SRC tasks. In this endeavor, the IBLT helps to explain how the cognitive processes are used, how the SRC task and Simon task become automatic, how the effects are attenuated when the two tasks are intermixed, and the effects for novel mixing situations. We compared the IBLT model predictions and fits to the human data for sequential effects as a function of whether the spatial mapping was compatible or incompatible, mapping repeats or switches, and when Simon or SRC task repeats or switches in the mixed Simon and SRC tasks.

A third project reported in Gonzalez, Dutt, Healy, Young, and Bourne (2009) presents a comparison of an IBLT model to a Strategy-Based Learning (SBL) model in a common task: dynamic visual detection. The SBL approach, when implemented in ACT-R, provides an account of human learning due to the use of a finite set of strategies (as opposed to the IBLT approach, which uses retrieval from memory and declarative knowledge from memory). We compared the two models based on (a) how well each model fits human learning data in the task; and (b) how well each model is able to reproduce the way humans, having learned in one scenario of the task, behave in a testing condition, where the scenarios are similar to or different from the training condition.

Taken together, these studies suggest that the IBLT presents an accurate and robust representation of the learning process in several diverse tasks. Because the IBLT has also shown accurate representations in many other tasks (Gonzalez et al., 2003; Gonzalez & Lebiere, 2005; Lejarraga, Dutt, & Gonzalez, 2010), we conclude that the theory is more general than it was initially conceived to be. The results generalize the IBLT's domain and application and show that it is well suited for other non-decision making tasks, such as the simple visual attention and search tasks summarized here. This ability is illustrated by the precision of the model’s predictions in several of the projects we have described.

The IBLT modeling tool, which was used in the MURI, is summarized in a technical report (Gonzalez, 2010).

2. IMPRINT Models

a. Problem studied. The human performance modeling tool used for this project is the Improved Performance Research Integration Tool (IMPRINT). IMPRINT has before now been mainly used for large-scale modeling. In the present project, the goal was to begin to develop the relationship between training variables and Soldier performance based on smaller-scale cognitive tasks, which before this project had not been done due to a lack of empirical data. In the present project, we used the data gathered in several cognitive experiments from our laboratory to help in understanding the use of IMPRINT for cognitive-level modeling, we collected information that could be used to inform the creation of a task taxonomy, and we learned how various aspects of training seen in the experimental setting could be implemented in IMPRINT. Specifically, three very different tasks--one a simple cognitive task, the second a task not only more complex and army-relevant but also involving a secondary or distractor task, and the third simulating part of a networked battlefield--were chosen to be modeled. The

goal of the modeling was to predict performance that reflected underlying cognitive processes as revealed by the experimental data in these three tasks.

b. Important results. The IMPRINT modeling part of the project simulated the experimental results from three very different cognitive experiments conducted in our laboratory. The three tasks were (a) digit data entry, a simple number typing task (Healy, Kole, Buck-Gengler, & Bourne, 2004); (b) the RADAR task developed at Carnegie Mellon University, which involved visual search and detection of targets among distractors (Young et al., in press); and (c) the information integration (fusion) task, developed to test memory for serially presented targets used to make firing decisions maximizing target damage (Ketels et al., 2010).

For digit data entry, we modeled Experiments 1 and 2 of Healy et al. (2004). Some of the specific aspects that were modeled were the contrast between repeated (Experiment 1) and non-repeated (Experiment 2) stimuli; the effects of changing hands in Experiment 2 (and what it means to use one's non-preferred hand); the common finding of chunking, in which the response times (RTs) for the third digit are longer than those for the second and fourth digits of the four-digit numbers; and improvement in RT along with deterioration in accuracy across trials. In the process of conducting this modeling, we broke down the responses of the subjects into their component parts and were able to determine differences for cognitive and motoric processing. Even more importantly, this was the first time any digit data entry responses had been examined so thoroughly at the individual subject level. Thus, we learned that chunking was in fact not universal across subjects, but rather represented a strategy choice. This choice was then successfully built into the model.

For RADAR, we modeled Experiment 1 of Young et al. (in press). The RADAR task was a more difficult task to model, as it had a simultaneous secondary task in some conditions. Specifically modeled were the secondary task's effect on the primary task performance at the time of the task and also the effect of the secondary task at training on performance at test, as well as the impact on performance (RT, hit rate, and false alarm rate) of the complex interaction between two variables that affected task difficulty--mapping type and processing load. For the purpose of modeling, the data needed to be reanalyzed at the most basic frame level, and in that reanalysis, interesting and complex patterns of learning (improvement), at least for the false alarm rate, were discovered. Specifically, there was an underlying low-level improvement throughout the session, but higher levels of improvement occurred when the task was most difficult in some way: either in the first block of each session, where the task was (relatively) new, or in the most difficult blocks, where the load was high and there was varied mapping (foils and targets of the same type).

Finally, the fusion task, developed relatively recently in our laboratory, was intended to be closer to a task that might occur in real army situations: where a number of targets are shown sequentially and then the subject must choose the best location that will damage the most targets according to various algorithms. Experiments 2 and 3 of Ketels et al. (2010) were chosen for modeling. The only difference between these two

experiments was the order of recalling the seven target locations and the firing decision. The observed firing decision did not differ very much between the two experiments, but the observed recall rate in Experiment 3, in which recall was done first, was far better than that of Experiment 2, in which recall followed the decision. Also, the observed firing location, in relation to the seven target locations, bore some resemblance to the serial recall curves. Thus, Experiment 3 was taken as the base experiment, with the idea that the amount recalled in both experiments should inform the firing decision but that the firing decision in Experiment 2 hurt memory for the seven locations, thus depressing the recall accuracy in Experiment 2 relative to Experiment 3. The Start-End Model (Henson, 1998) was found to be most useful as a starting point for understanding serial recall curves. Using an abbreviated form of that model, the recall and firing decision results were both modeled successfully in IMPRINT.

The codes for these models are available on compact disk upon request from Alice Healy.

3. Model Comparison and Evaluation

a. Problem studied. In the original proposal, the last of the three "Technical Approach" items was "III. Explication of two different modeling approaches," in turn separated in three parts: A. Modeling with IMPRINT, B. Modeling with ACT-R, and C. Mathematical soundness and computational feasibility of modeling efforts. This last item was investigated as a team effort led by co-PI Bengt Fornberg. The effort started as planned by comparing the effectiveness and computational speeds of separately developed IMPRINT and ACT-R models of laboratory experiments involving two different tasks. However, the discovery of major unexpected opportunities in terms of speedup and scalability when translating the IMPRINT models to a language optimized for scientific computing (Matlab) caused us to extend the scope of the model evaluation task to also include an extensive study of the additional opportunities this translation provided, especially in terms of parameter optimization.

b. Important results. Several comparisons and evaluations of IMPRINT and ACT-R models of two laboratory tasks (keystroke data entry and RADAR visual search) were conducted during the MURI. The first step in our MURI effort on this issue was to critically assess the relative performance (in terms of accuracy, speed, and coding complexity) of the IMPRINT and the ACT-R models that were developed by two different specialized expert teams, one for each platform. A critical issue that was explored was the issue of scalability: the feasibility of efficiently scaling up models towards much larger future problem sizes and complexities. Although large models on these platforms had previously been implemented and run for extended times on giant computer systems, this fact was as likely to raise concerns as it was to alleviate concerns about future possibilities of scaling models upward.

After the IMPRINT and ACT-R teams had produced well-fit models for the two tasks, it become clear that, in spite of the great conceptual differences between the modeling platforms, both accuracy and computational cost were comparable (when using

similar hardware). From a military perspective, key distinguishing factors between the two platforms would rather be other capabilities, such as the ability to interface to other systems.

Performance comparisons between altogether different computational systems were missing in the previous literature, but are necessary in order to get an independent objective point of reference. Much of our effort, therefore, became directed towards exploring whether the equation-based approach in our IMPRINT models could be substantially speeded up by being reprogrammed in Matlab (one of many highly effective scientific languages in very widespread use, with C++ and Fortran being two other options). Matlab was selected here mainly for its great ease of use, and its very effective and convenient capabilities for porting from standard PC-based platforms to parallel and distributed processing environments. The result of this language translation was so encouraging that the original direction of the Model Evaluation effort was promptly supplemented by an additional effort of exploring how the speedup using Matlab of the order of 10,000 (on comparable hardware) could best be utilized to provide additional modeling capabilities. A particularly promising opportunity that was pursued concerned automated optimization of model parameters through the use of global optimization algorithms, such as simulated annealing and genetic algorithms. It was also discovered that Radial basis functions (RBFs) offer major additional opportunities for speeding up the evaluation of models and for interactive visualization of multivariate data sets, including the optimized parameter spaces of the models.

Because further speedup factors of several orders of magnitude can readily be achieved by using parallel or distributed computing, the scalability is no longer as uncertain as it was perceived to be when the present MURI was initiated.

The details of the model comparison component of the MURI are summarized in a technical report (see Fornberg, Raymond, Buck-Gengler, Healy, & Bourne, 2010).

D. Summary

During the five years of the Training MURI (5/1/05-9/30/10), significant progress was made on all three components of the project: experiments, taxonomy, and models. New experiments were conducted on (a) the development and testing of training principles, (b) the acquisition and retention of basic components of skill, and (c) training effects associated with levels of automation, individual differences, and team performance. To render the study of training effects tractable and to guide research, we developed a multi-dimensional taxonomy, which provides a framework by which training effects can be assessed and predicted for any task. The taxonomy involves a four-dimensional decomposition of the training space and includes separate dimensions of classification for task description, training procedure, and the context and assessment of task performance. The training principles are considered the fourth dimension. The component of the project devoted to models consists of three parts. The work on ACT-R developed models of the simple data entry task, of the more complex RADAR task, and of stimulus-response compatibility effects. It also involved development of a Visual

Basic modeling tool. The work on IMPRINT developed a model of data entry, a model of the RADAR task, and a model of information integration (fusion). The part on model assessment focused on model optimization. The Matlab platform and the algorithms included in the IMPRINT models were used for this purpose. These various efforts yielded many submitted manuscripts, peer-reviewed journal publications, chapters published in books or conference proceedings, presentations at professional meetings, master's theses, and doctoral dissertations. We also pursued numerous points of transition between the results of basic research in this project and the eventual applied needs of Army trainers, including the specification of performance shaping functions, which are quantitative versions of training principles that can be incorporated into IMPRINT.

E. Bibliography

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Anderson, L. S., Healy, A. F., Kole, J. A., & Bourne, L. E., Jr. (2010, April). *The clicker technique: An effective method of teaching compression*. Poster presented at the 80th Annual Convention of the Rocky Mountain Psychological Association, Denver, CO.

Bonk, W. J., & Healy, A. F. (2010). Learning and memory for sequences of pictures, words, and spatial locations: An exploration of serial position effects. *American Journal of Psychology, 123,* 137-168.

Bourne, L. E., Jr., Raymond, W. D., & Healy, A. F. *Quantifying performance effects of training manipulations: Performance shaping functions based on selected training principles*. (Technical Report). University of Colorado, Boulder, Center for Research on Training.

Clegg, B. A., & Heggestad, E. D. (2010). *Automation and effective training: Training for the networked battlefield MURI project*. (Technical Report). University of Colorado, Boulder, Center for Research on Training.

Dutt, V., Gonzalez, C., Yamaguchi, M., & Proctor, R. W. (2010). *Instance-based learning models of SRC and Simon effects*. Unpublished manuscript in preparation.

Dutt, V., Yamaguchi, M., Gonzalez, C., & Proctor, R. W. (2009). An instance-based learning model of stimulus-response compatibility effects in mixed location-relevant and location-irrelevant tasks. In A. Howes, D. Peebles & R. Cooper (Eds.), *9th International Conference on Cognitive Modeling – ICCM2009*. Manchester, UK.

Fornberg, B., Raymond, W. D., Buck-Gengler, C. J., Healy, A. F., & Bourne, L. E., Jr. (2010). *Model evaluation: ACT-R, IMPRINT, and Matlab: Relative performances and new opportunities*. (Technical Report). University of Colorado, Boulder, Center for Research on Training.

Fu, W., Gonzalez, C., Healy, A. F., Kole, J. A., & Bourne, L. E., Jr. (2006). Building predictive models of skill acquisition in a data entry task. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting (HFES 50th Annual Meeting)* (pp. 1122-1126). Santa Monica, CA: Human Factors and Ergonomics Society.

Gonzalez, C. (2010). *The instance-based learning modeling tool*. (Technical Report). University of Colorado, Boulder, Center for Research on Training.

Gonzalez, C., Best, B. J., Healy, A. F., Bourne, L. E., Jr, & Kole, J. A. (2010). *A cognitive modeling account of fatigue*. Manuscript submitted for publication.

Gonzalez, C., Dutt, V., Healy, A. F., Young, M. D., & Bourne, L. E. (2009). Comparison of instance and strategy models in ACT-R. In A. Howes, D. Peebles & R. Cooper (Eds.), *9th International Conference on Cognitive Modeling – ICCM2009*. Manchester, UK.

Gonzalez, C., Fu, W., Healy, A. F., Kole, J. A., & Bourne, L. E., Jr (2006). *ACT-R models of training data entry skills. In Proceedings of the Conference on Behavior Representation in Modeling and Simulation (BRIMS 2006)* (pp. 101-109). Baltimore, MD.

Gonzalez, C., & Lebiere, C. (2005). Instance-based cognitive models of decision making. In D. Zizzo & A. Courakis (Eds.), *Transfer of knowledge in economic decision-making* (pp. 148-165). New York: Macmillan (Palgrave Macmillan).

Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science, 27,* 591-635.

Healy, A. F., & Bourne, L. E., Jr. (Eds.). (1995). *Learning and memory of knowledge and skills: Durability and specificity*. Thousand Oaks, CA: Sage.

Healy, A. F., & Bourne, L. E., Jr. (2010). *Experimental research on training principles: MURI Research at the University of Colorado*. (Technical Report). University of Colorado, Boulder, Center for Research on Training.

Healy, A. F., Clawson, D. M., McNamara, D. S., Marmie, W. R., Schneider, V. I., Rickard, T. C., Crutcher, R. J., King, C., Ericsson, K. A., & Bourne, L. E., Jr. (1993). The long-term retention of knowledge and skills. In D. Medin (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 30, pp. 135-164). New York: Academic Press.

Healy, A. F., Kole, J. A., Buck-Gengler, C. J., & Bourne, L. E., Jr. (2004). Effects of prolonged work on data entry speed and accuracy. *Journal of Experimental Psychology: Applied, 10,* 188-199.

Healy, A. F., Kole, J. A., Wohldmann, E. L., Buck-Gengler, C. J., & Bourne, L. E., Jr. (in press). Data entry: A window to principles of training. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: A festschrift in honor of Robert A. Bjork*. New York: Psychology Press.

Healy, A. F., Schneider, V. I., & Bourne, L. E., Jr. (2010). *Towards the improvement of astronaut training: A literature review of empirical evidence for training principles*. (Technical Report). University of Colorado, Boulder, Center for Research on Training.

Healy, A. F., Wohldmann, E. L., Kole, J. A., Schneider, V. I., Shea, K. M., & Bourne, L. E., Jr. (in press). Training for efficient, durable, and flexible performance in the military. In W. Arthur, Jr., E. A. Day, W. Bennett, Jr., & A. Portrey (Eds.), *Individual and team skill decay: State of the science and implications for practice*. New York: Taylor & Francis.

Henson, R. N. (1998). Short-term memory for serial order: The Start-End Model. *Cognitive Psychology, 36*, 73-137.

Ketels, S. L., Healy, A. F., Wickens, C. D., Buck-Gengler, C. J., & Bourne, L. E., Jr. (2010, April). *Spatial list learning and decision making in the fusion paradigm*. Poster presented at the 80th Annual Convention of the Rocky Mountain Psychological Association, Denver, CO.

Kole, J. A., & Healy, A. F. (2007). Using prior knowledge to minimize interference when learning large amounts of information. *Memory & Cognition*, *35*, 124-137.

Kole, J. A., & Healy, A. F. (2008, November). *Retrieval-induced forgetting: Examining representation weakening and retrieval competition accounts*. Poster presented at the 49th Annual Meeting of the Psychonomic Society, Chicago, IL.

Kole, J. A., & Healy, A. F. (2010, April). *Memory for facts about people: Familiarity, relatedness, degree of genetic similarity, and gender congruency*. Poster presented at the 80th Annual Convention of the Rocky Mountain Psychological Association, Denver, CO.

Kole, J. A., Healy, A. F., & Bourne, L. E. (2008). Cognitive complications moderate the speed-accuracy tradeoff in data entry: A cognitive antidote to inhibition. *Applied Cognitive Psychology, 22,* 917-937.

Kole, J. A., Healy, A. F., Fierman, D. M., & Bourne, L. E., Jr. (2010). Contextual memory and skill transfer in category search. *Memory & Cognition, 38,* 67-82.

Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology, 78,* 311-328.

Lejarraga, T., Dutt, V., & Gonzalez, C. (2010). *Instance-based learning in repeated binary choice*. Unpublished manuscript under review.

Lohse, K. R., Healy, A. F., & Sherwood, D. E. (2009, November). *Task-level and effector-level representations in intermanual transfer of motor skills*. Poster presented at the 50th Annual Meeting of the Psychonomic Society, Boston, MA.

Proctor, R. W., Yamaguchi, M., & Miles, J. D. (2010). *Acquisition and transfer of basic skill components*. (Technical Report). University of Colorado, Boulder, Center for Research on Training.

Proctor, R. W., Yamaguchi, M., Zhang, Y., & Vu, K. L. (2009). Influence of visual stimulus mode on transfer of acquired spatial associations. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 35,* 434-445.

Raymond, W. D., Healy, A. F., & Bourne, L. E., Jr. (2010). *The MURI training taxonomy*. (Technical Report). University of Colorado, Boulder, Center for Research on Training.

Roth, J. T. (1992). *Reliability and validity assessment of a taxonomy for predicting relative stressor effects on human task performance*. Technical Report 5060-1 prepared under contract DNA001-90-C-0139, Boulder, CO: Micro Analysis and Design, Inc.

Schneider, V. I., Healy, A. F., Buck-Gengler, C. J., Barshi, I., & Bourne, L. E., Jr. (2008, November). *Effects of presenting navigation instructions twice in the same or different modalities*. Paper presented at the 49th Annual Meeting of the Psychonomic Society, Chicago, IL.

Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology, 41,* 1-19.

Staal, M. A., Bolton, A. E., Yaroush, R. A., & Bourne, L. E., Jr. (2008). Cognitive performance and resilience to stress. In B. Lukey & V. Tepe (Eds). *Biobehavioral resilience to stress* (pp. 259-300). London: Francis & Taylor.

Welford, A. T. (1976). *Skilled performance: Perceptual & motor skills*. Englewood Cliffs, NJ: Prentice Hall.

Wickens, C. D., & Holland, J. G. (2000). Memory, learning, and training. In C. D. Wickens & J. G. Holland (Eds.), *Engineering psychology and human performance* (3rd ed., pp. 241-284). New York: Harper Collins.

Wickens, C. D., Ketels, S. L., Healy, A. F., Buck-Gengler, C. J., & Bourne, L. E., Jr. (in press). The anchoring heuristic in intelligence integration: A bias in need of de-biasing. *Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.

Wohldmann, E. L., & Healy, A. F. (2010). Exploring specificity of speeded aiming movements: Examining different measures of transfer. *Memory & Cognition, 38,* 344-355.

Wohldmann, E. L., Healy, A. F., & Bourne, L. E., Jr. (2007). Pushing the limits of imagination: Mental practice for learning sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 254-261.

Wohldmann, E. L., Healy, A. F., & Bourne, L. E., Jr. (2008a). A mental practice superiority effect: Less retroactive interference and more transfer than physical practice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 823-833.

Wohldmann, E. L., Healy, A. F., & Bourne, L. E., Jr. (2008b). Global inhibition and midcourse corrections in speeded aiming. *Memory & Cognition, 36,* 1228-1235.

Wohldmann, E. L., Healy, A. F., & Bourne, L. E., Jr. (2010). Task integration in time production. *Attention, Perception, & Psychophysics, 72*, 1130-1143.

Yamaguchi, M., Dutt, V., Gonzalez, C., & Proctor, R. W. (2010). *Cognitive mechanisms of the stimulus-response compatibility effects in mixed location-relevant and location-irrelevant tasks*. Manuscript submitted for publication.

Yamaguchi, M., & Proctor, R. W. (2006). Stimulus-response compatibility with pure and mixed mappings in a flight task environment. *Journal of Experimental Psychology: Applied, 12*, 207-222.

Young, M. D. (2010). *Artificial grammar learning: Implicit and explicit components for retention and transfer*. Unpublished doctoral dissertation, University of Colorado, Boulder.

Young, M. D., Healy, A. F., Gonzalez, C., Dutt, V., & Bourne, L. E., Jr. (in press). Effects of training with added difficulties on RADAR detection. *Applied Cognitive Psychology*.

Quantifying Performance Effects of Training Manipulations: Performance Shaping Functions Based on Selected Training Principles

Lyle E. Bourne, Jr., William D. Raymond, and Alice F. Healy

University of Colorado, Boulder

Quantifying Performance Effects of Training Manipulations:
Performance Shaping Functions Based on Selected Training Principles
Lyle E. Bourne, Jr., William D. Raymond, and Alice F. Healy

Training principles that we have identified and supported empirically can be expressed as equations, and these equations, in turn, can be incorporated into IMPRINT for purposes of predicting post-training performance.

There are two fundamental principles of training that derive from the work of others and that were confirmed in our research:

(A) Practice improves performance (power law of practice)

(1) $$p = a + bN^{-c}$$

where p is performance (e.g., RT or errors), N is number of practice trials, a is asymptotic performance, c is rate of learning, and b is a scaling parameter.

Applicable to the following IMPRINT task taxons: numerical analysis, information processing, fine motor discrete, fine motor continuous, gross motor – light

(B) Power law of forgetting

(2) $$p = d + eT^{-f}$$

where p is performance, T is time since learning (retention interval), d is the degree of learning, f is the rate of forgetting, and e is a scaling parameter.

Applicable to the following IMPRINT task taxons: numerical analysis, information processing, fine motor discrete, fine motor continuous, gross motor – light

There are additional, more specific training principles that were formulated during the course of the MURI research. The following are four examples:

(C) Deep processing (levels of processing)

Deep processing during training improves performance after training

(3) $$p_i = g_i p_n$$

where p_i is performance (RT or errors) after training following a deep processing condition i during training, p_n is performance after training following the most shallow processing requirement during training, and g_i (< 1) is the benefit from a deep processing condition i

Applicable to the following IMPRINT task taxons: numerical analysis, information processing

(D) Generation (the generation effect)

Subject generation of items (as opposed to item reading) during training improves performance after training.

(4) $$p_i = h_i p_n$$

where p_i is performance (RT or errors) after training following a deep processing condition i during training, p_n is performance after training following the most shallow processing requirement during training, and h_i (< 1) is the benefit from a deep processing condition i

Applicable to the following IMPRINT task taxons: numerical analysis, information processing

(E) Difficulty

Difficulty (e.g., contextual interference) during learning lowers accuracy (increases errors) during training but improves long-term retention.

After training,

(5) $$p_1 = (1+k)p_n$$

where p_1 is performance (proportion of errors) during training under contextual interference, p_n is performance during training under no interference conditions, and k ($-1<k<0$) is the magnitude of the interference effect at training.

After a delay,

(6) $$p_2 = (1+q)p_m$$

where p_2 is performance after a delay following contextual interference during training, p_m is performance after a delay under no interference conditions during training, and q ($0<q<1$) is the magnitude of the interference effect after a delay.

Applicable to the following IMPRINT task taxons: numerical analysis, information processing, fine motor discrete, fine motor continuous

(F) Mnemonic procedures

One type of effective mnemonic procedure that involves relating facts to be learned to already well-known facts during training improves performance after training as well as after a delay (i.e., strategic use of knowledge principle).

At the end of training,

(7) $$p_1 = (1+r)p_n$$

where p_1 is performance (in this case, proportion of correct responses) at the end of training following a mnemonic procedure condition during training, p_n is performance after training following no mnemonic processing requirement during training, and r (> 0) is the benefit from strategic use of knowledge.

After a delay following training,

$$p_2 = (1+s)p_m \tag{8}$$

where p_2 is performance after a retention interval following training with a mnemonic procedure condition, p_m is performance after a retention interval following training with no mnemonic processing requirement, and s (> 0) is the benefit from strategic use of knowledge.

Applicable to the following IMPRINT task taxons: numerical analysis, information processing but not fine motor discrete

Illustrative Applications

We looked at two manipulations, one involving training difficulty and the second involving mnemonic procedures. We chose the first because of its striking, unintuitive results. Specifically, training under difficult conditions led to worse performance at the end of training but better performance after a 1-week delay. We chose the second because of its massive positive effects both immediately after training and after a 1-week delay.

(E) Difficulty

The difficulty manipulation is based on an assessment involving the direction of associations (i.e., for a translation task, the easier French-to-English translation direction is compared to the harder English-to-French direction). Data from Schneider, Healy, and Bourne (2002) were used to derive the following table:

Effect of Translation Direction on Accuracy

Training type	*End of Training (3 repetitions)*	*Retention (in both directions across participants)*
Easy (French to English)	0%	0%
Hard (English to French)	-37%	+23%

Note that the difficulty manipulation used here hurt performance at the end of training but, despite the lower amount learned during training, aided performance at the retention test. These numbers, based on proportion of correct translation responses, were derived from tests given immediately after training and then again after a 1-week delay. The tests given at the end of training were restricted to the translation direction used during training, whereas the retention

tests given 1 week later occurred in both directions (across participants). The easy translation direction is used as a baseline (i.e., set to 0% separately for both the training and the retention test) to assess the magnitude and direction of the effect of translation direction on both training and retention. There was no intermediate level of training difficulty in this experiment, although we might make the reasonable guess that performance with an intermediate difficulty could be derived by interpolation. There was only a single retention interval in the present study (1 week), but we assume that the forgetting function would not interact with the delay. In any event, because of the procedure used to set the baseline separately for both training and retention, the percentages do not reflect the forgetting that occurred over the 1-week retention interval.

Recalling the relevant equations:

(5) $$p_1 = (1+k)p_n$$

(6) $$p_2 = (1+q)p_m$$

Thus, for Equations 5 and 6, we estimate that $k = -.37$ and $q = .23$.

(F) Mnemonic procedures

The mnemonic manipulation is based on a comparison of a situation in which new information is learned about items for which there is high prior knowledge (i.e., friends or relatives) with a situation in which the same new information is learned instead about items for which there is no prior knowledge (i.e., unfamiliar individuals). Recently collected data, following up the published reference by Kole and Healy (2007), were used to derive the following table:

In the second

Effect of Strategic Use of Prior Knowledge on Accuracy

Training type	*End of Training*	*Retention*
Low Knowledge	0%	0%
High Knowledge	+337%	+184%

These numbers, based on proportion of correct associative recall responses, were derived from tests given immediately after training and then again after a 1-week retention interval. Equivalent tests were given at the two times. The low-knowledge training condition is used as a baseline (i.e., set to 0 separately for both the immediate test and the delayed test) to assess the magnitude and direction of the effect of using a mnemonic procedure based on prior knowledge on both training and retention. There was no intermediate level of degree of prior knowledge in this experiment, although we might make the reasonable guess that performance with an intermediate difficulty could be derived by interpolation. There was only a single retention interval in the present study (1 week), but we assume that the forgetting function would not interact with the delay. In any event, because of the procedure used to set the baseline separately for both the immediate and delayed test, the percentages do not reflect the forgetting that occurred over the 1-week retention interval.

Recalling the relevant equations:

(7) $$p_1 = (1+r)p_n$$

(8) $$p_2 = (1+s)p_m$$

Thus, for Equations 7 and 8, we estimate that $r = 3.37$ and $s = 1.84$.

References

Kole, J. A., & Healy, A. F. (2007). Using prior knowledge to minimize interference when learning large amounts of information. *Memory & Cognition, 35*, 124-137

Schneider, V. I., Healy, A. F., & Bourne, L. E., Jr. (2002). What is learned under difficult conditions is hard to forget: Contextual interference effects in foreign vocabulary acquisition, retention, and transfer. *Journal of Memory and Language*, *46*, 419-440.

Automation and Effective Training

Training for the Networked Battlefield
MURI Project

Army Research Office Grant W911NF-05-1-0153 (PI: Alice Healy)

Benjamin A. Clegg
Colorado State University
&
Eric D. Heggestad
University of North Carolina - Charlotte

Benjamin Clegg & Eric Heggestad

EXECUTIVE SUMMARY

Automation may seem like it is helping you, especially if you need more help, but it is not helping you learn.

Our research on the role of automation in training offers the following core findings:

- Automation frequently has a negative impact on training
    - Find negative effects of automation in both a microworld simulation situation and in an operationally relevant Predator UAV simulator
    - Not all levels of automation have equivalent costs for learning
    - The costs of automation for learning merits examination in other tasks and environments

- Slowly taking away the availability of automation over time was insufficient to avoid the negative consequences on learning
    - Training strategies to effectively overcome the costs of automation need to be developed

- There were specific situations where automation was beneficial, but these benefits did not extend to the underlying learning
    - There was a benefit to performance early in training, particularly for lower aptitude trainees

- The pattern of results most closely matches the Specificity of Training Principle, one of the core learning and training principles within this MURI. The findings suggest a new range of domains to which this principle might productively be applied.

AUTOMATION AND TRAINING

The increasing presence of automation is changing the nature of a wide variety of tasks. Tele-operations and robotic systems are likely to play an increasingly important role within the future networked battlefield. Automation offers the potential to make tasks easier. This might allow a single operator to control more systems, and be more productive; or might allow individuals without highly specialized training to accomplish complex tasks. This begs the question of how best to train operators to run such systems. Additionally, automation can function as a training aid – supporting novice operators, and perhaps even making the task more manageable, allowing them to focus on learning.

Although the scientific knowledge on what automation is and how it can aid human operators is flourishing, there is still much to learn about the role of automation in training. We found no previous research that has looked at how operator individual differences might relate successful incorporation of automation into training programs. Our research has sought to address this important issue.

Specifically, the key goals and objectives guiding our research were:

- To examine the role of automation in skill learning
- To determine whether the aptitude of the learner interacts with presence of automation to influence the effectiveness of training

Levels of Automation

Instead of regarding automation as a binary option (“automation” vs. “no automation”), Sheridan and Verplank (1978; see also Endsley & Kaber, 1999; Kaber & Endsley, 2004; Kaber, Onal, & Endsley, 2000; Parasuraman, Sheridan, & Wickens, 2000) proposed that allocation of function between human and machine spans a series of different possible levels of automation. Such advances offer a significant degree of flexibility in developing varieties of automation across a broad spectrum of tasks. One critical transition within these taxonomies occurs for the change in how automation is initiated. For example, the human operator can actively engage automation or he or she can veto power over automation that is automatically initiated. This distinction is sometimes referred to as “management by consent” versus “management by exception” (Billings, 1997).

While such developments in understanding levels of automation offer important insights into the different forms and types available, they also raise the question of how to instantiate automation during training to maximize efficiency, durability, and flexibility of learning. Training with highly automated systems is becoming increasingly common. However, operators might not even develop all the appropriate underlying skills and knowledge if they are only trained while relying on automated aides that support their performance (Moray, 1986).

Interactions Between Individual Differences and Training Effectiveness

Clamann, Wright, and Kaber (2002) highlight that problems adapting to automation seem more pronounced for cognitive tasks (analysis and decision aids) versus automation applied to lower level components (information acquisition and action implementation). Findings such as these raise the possibility that determining the effectiveness of training with these more demanding forms of automation will be influenced by individual differences between operators. Individual differences in cognitive abilities are related to variations in learning, retention, and transfer performance in training contexts. Ackerman (1988) showed that as learning progresses, the rate of skill acquisition relates to individual variation in different cognitive components. Those higher in general cognitive ability (‘g’) show superior performance during the initial, declarative knowledge phase of learning; then perceptual speed abilities are related to performance in the next phase, knowledge compilation; and, later in learning, psychomotor abilities are related to performance in the procedural phase.

Research from an aptitude treatment interaction (ATI) perspective (Cronbach, 1957) has indicated that relationships between cognitive abilities and training outcomes differ as a function of the nature of training. Variations in training have also been shown to interact with cognitive ability to influence performance in a transfer of training environment (Goska & Ackerman, 1996). Thus, it is clear that the effectiveness of a training intervention depends in part on the characteristics of the trainees.

Given the importance of interactions between individual differences in cognitive ability and forms of training on training outcomes, one imperative question within our research was to examine how aptitude and training type interact, with training type defined by levels of automation. While other research implies that intermediate levels of automation may prove most effective during training (e.g., Clamann, Wright, & Kaber, 2002; Clegg, Blalock, Rodriguez, & Moray, 2010) thus far such work has not systematically explored individual differences in performance, and in particular conducted an examination of potential aptitude-treatment interactions. That is to say, the effects of various types of automation on learner performance are likely to vary as a function of the traits of the individual.

Research Platforms

Our initial studies were conducted using a simulated process control task, “Pasteuriser”, developed from an earlier micro-world simulation used by Moray and others (e.g., Lee 1992; Lee & Moray, 1992; Muir, 1987; Muir & Moray, 1996; see also Reising & Sanderson, 2002). The properties of Pasteuriser are comparatively well known, including established knowledge about the amount of practice required on the task, the shape of the learning curves, etc. Complexity in the operator’s task arises from the interaction of three subsystems, plus the presence of competing goals, and also the dynamics that incorporate time lags (for more details on the simulation see Lee, 1992). The version developed by Moray, Rodriguez and Clegg (2000) allows a choice of level of automation under which the operator will run the system. In the higher levels of automation, three subsystems can be controlled either manually or automatically. This platform provided the foundational data for Colorado State’s contribution to the MURI project.

Subsequent studies were carried out using alternative platforms to extend the research into operational relevant domains and into team performance. Our approach was to look for a military-task simulation with which to examine the applicability of our findings. In seeking a dynamic task with a relative rapid initial learning curve, we were granted access to the Predator unmanned aerial vehicle (UAV) synthetic task environment (STE), developed at the Air Force Research Laboratory’s Warfighter Training Research Division (Martin, Lyon, & Schreiber, 1998). The platform was developed to assess the acquisition of key skills required of UAV pilots, and thus has value for us in terms of both its relation to this real military task and its close correspondence to a wide variety of current and future military tasks. A further advantage of the use of this platform was the presence of structured training, in contrast to the trial-and-error training in the Pasteuriser task, adding further scope for generalization of our previous findings

The UAV STE program was not originally conceived as a platform to assess the role of automation in training. However, the design of the basic maneuvering modules, intended to teach control of the Predator, incorporated automation to allow for part-task training (of a type very similar to that which we included in the study reported below). By changing the structure of the modules, we were therefore able to assess whether the ability to use automation to focus learning on specific aspects of the task improves learning, or whether, as in our previous studies, the result is impoverished learning of the system compared to individuals learning with no such automation-based support.

FINDINGS

Taken together the results across our experiments illustrate the impact of automation on training and serve to highlight the possible deleterious effects of the inclusion of automation within training. Our data offer unique insights, such as providing the first evidence of an aptitude-automation interaction in training effectiveness.

In the initial experiment (Clegg, Heggestad, & Blalock, 2010), training was conducted with operators either performing with no automation present (manual control), with short-term assistance from automation that required active engagement by the user (user-initiated automation), or with short-term assistance engaged by the automation unless vetoed by the user (automatically-initiated automation). Data were collected from more than 350 participants, each of who completed a cognitive ability battery and 2½ hours of training on the Pasteuriser task.

Figures 1 and 2 present the results for two performance metrics for each training condition over trials of learning. As shown, there was a benefit of training with automation early in training. More specifically, individuals in the manual control (no automation) group performed less well than participants in the two automation conditions in the first learning trial. These benefits, however, rapidly diminished and ultimately even reversed with an advantage for operators trained solely through manual control (see Figure 2).

Our results also offer evidence consistent with the notion that an automation by consent approach is generally preferable to one of management by exception (see Liu, Wasson & Vincenzi, 2009; Ruff, Nayarana, & Draper, 2002; but see Olson & Sarter, 2001). Within our initial experiment, decrements in the development of underlying knowledge (seen when automation was removed) were only apparent in the case of automatically-initiated automation (see Figure 3).

Figure 1
Good juice production across training for Manual Control (MC), User-initiated automation (UIA), and Automatically-initiated automation (AIA) groups in the Pasteuriser task. Good juice results from operator putting simulation in the desired state, and these data show initial benefits from automation (block 1), but no differences across groups with increased training.

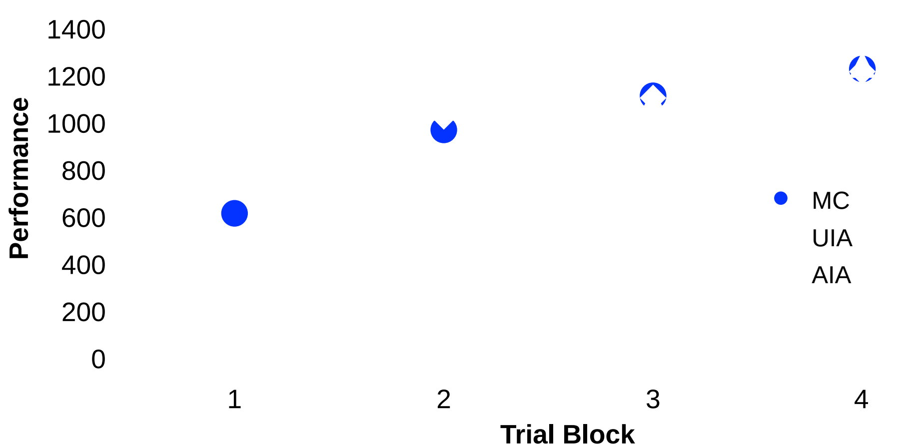


Figure 2
Spoiled juice across training for Manual Control (MC), User-initiated automation (UIA), and Automatically-initiated automation (AIA) groups in the Pasteuriser task. These data show initial benefits from automation (block 1), but superior performance from the group trained without automation by the end of training (block 4).

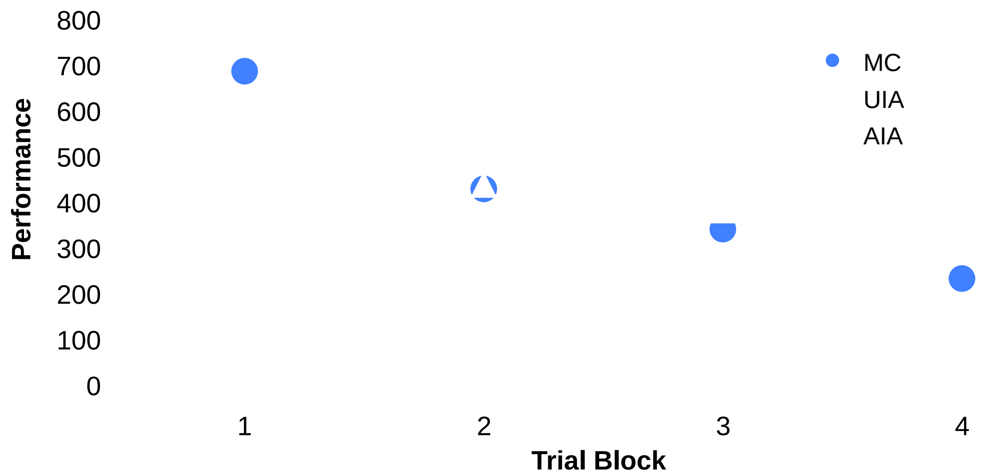
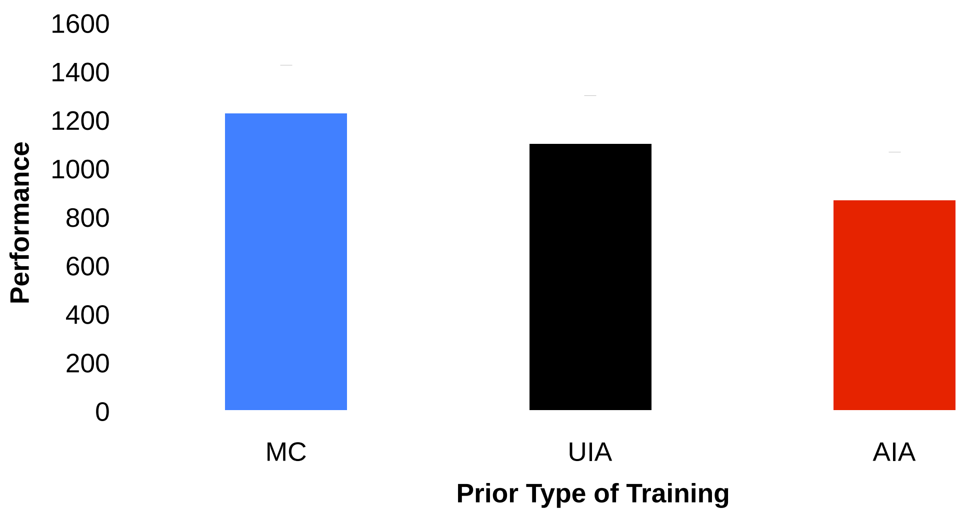


Figure 3
Overall juice production (good juice production minus bad juice production) in the Pasteuriser task with automation removed, as a function of type of prior training. Prior training comprised Manual Control (MC), User-initiated automation (UIA), and Automatically-initiated automation (AIA) groups.



Individual difference measures were collected using Educational Test Service's Kit of Factor-Referenced Cognitive tests. The correlations between the specific abilities measured and early and late performance are presented in Tables 1 and 2 for each of the experimental conditions. As shown, stronger correlations were observed within the MC condition than in conditions involving performance with automation.

Table 1
Correlations between abilities and Block 1 performance in Pasteuriser as a function of type of automation.

	Manual Control	User Activated	Auto. Activated
Reasoning	**.21**	-.13	**.25**
Quantitative	**.44**	.01	.16
Verbal	.02	-.11	.09
Spatial	**.21**	.13	.14
Perceptual Speed	-.03	-.15	-.16
g	**.40**	.03	**.24**

Note: Values shown in **boldface** are statistically significantly different from zero.

Table 2
Correlations between abilities and Block 4 performance in Pasteuriser as a function of type of automation.

	Manual Control	User Activated	Auto. Activated
Reasoning	**.19**	-.04	.10
Quantitative	**.29**	.12	**.20**
Verbal	.07	-.09	.17
Spatial	**.18**	.06	.05
Perceptual Speed	-.06	.00	-.04
g	**.30**	.05	**.18**

Note: Values shown in **boldface** are statistically significantly different from zero.

A moderated regression analysis with good juice production in Block 1 as the dependent variable was conducted. Predictors for the model comprised g, two dummy variables representing training condition (with MC representing the base group), and two interaction terms. g was expected to be a significant predictor, indicating a general relationship between g and good juice production across the three conditions. More importantly, we expected that the interaction terms would also be statistically significant, with negative betas. Given that the MC condition was chosen as the base group, negative beta coefficients would indicate the relationship between g and performance is less strong in the two automation groups than in the MC group.

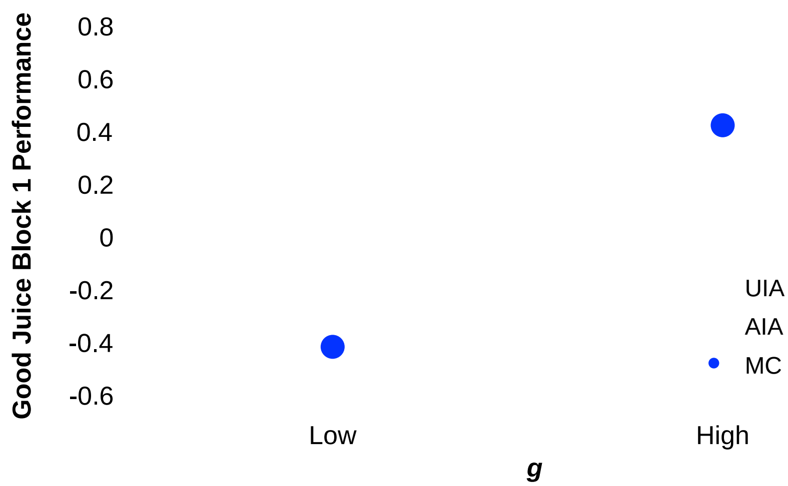
The results of the regression analysis are presented in Table 3. Significant positive coefficients for UIA and AIA indicate superior performance to MC. The beta coefficient for g was significant and positive, and the beta coefficients for the two interaction terms were significant and negative. A graphic representation of these results is presented in Figure 4. The figure reveals that higher g trainees perform equivalently across types of training. However, a pronounced difference was seen among lower g trainees; lower g trainees in the two automation conditions performed better than lower g trainees in the MC condition.

Table 3.
Moderated Regression Analyses on good juice production scores in Pasteuriser.

Predictor	Block 1	Block 4
User Activated Automation (dummy)	0.33**	0.03
Computer Activated Automation (dummy)	0.40**	0.01
g	0.42**	0.35**
Interaction for User Activated	-0.22**	-0.20**
Interaction for User Activated	-0.11	-0.11

Note. Values in the table are standardized beta coefficients. * $p < .05$; ** $p < .01$.

Figure 4
Automation type by g interaction in the prediction of Good Juice production in the Pasteuriser task. Training comprised Manual Control (MC), User-initiated automation (UIA), and Automatically-initiated automation (AIA) groups.



Although effects of variations in automation on training in this particular task were generally small, and tended to decrease with ongoing practice, the presence of an interaction with aptitude suggests an important set of considerations for designing and instituting training with automation. Selecting levels of automation for use within systems clearly has implications for what operators will learn from their training, but these implications will vary with the aptitude of individuals.

Our second experiment began to explore ways in which the negative consequences of automation in training might be removed. Once again employing the Pasteuriser platform, we examined two different approaches: gradually reducing the number of subsystems controlled by automation (Decreasing Automation), and varying which subsystem the operator controlled to induce part-task training (Random Automation).

The data once again show negative effects on learning from the presence of automation in training (see Figure 5). Neither gradually removing automation, nor using automation to impose the need for an operator to learn the functioning of specific subsystems was effective.

Our final experiment in this series (Blitch & Clegg, 2010) examined the impact of automation during training on learning within the STE Predator UAV platform. After some basic familiarization, participants were trained either with manual control over all the flight systems or with automation assisting with pitch and throttle. Participants then completed a series of manual control trials, and then were required to perform a novel landing task. The data (see Figure 6) are consistent with the findings from the previous project experiments. The use of automation in training led to poorer acquisition of knowledge of how to control the UAV.

These data offer evidence that the type of effects observed in the previous microworld simulations, selected because they contain properties relevant to many operational tasks, can also be observed directly within the setting of military tasks.

Figure 5
Good juice production at the end of training for Manual Control, Decreasing Automation, and Random Automation in the Pasteuriser task. These data show that withdrawing automation or utilizing automation for part-task training resulted in significantly worse learning than found for operators trained with manual control over the system.

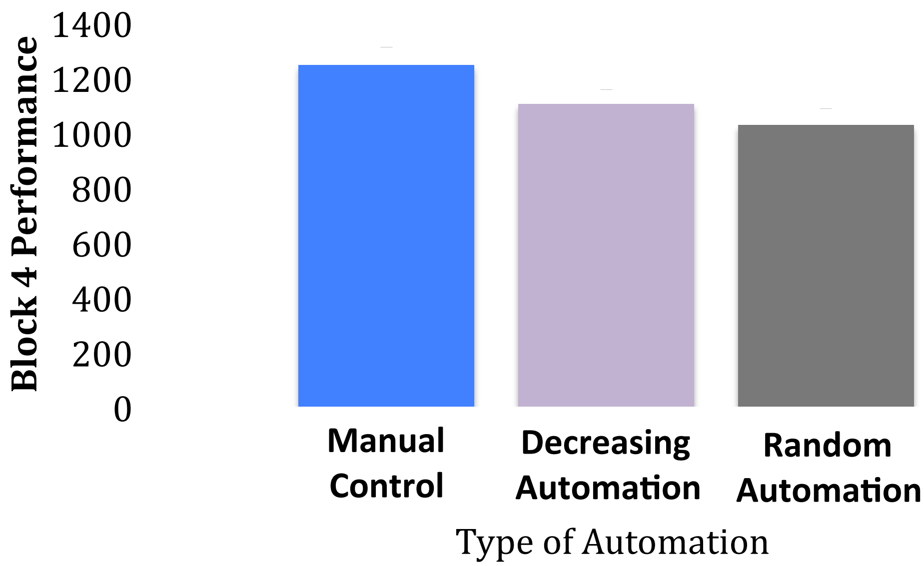
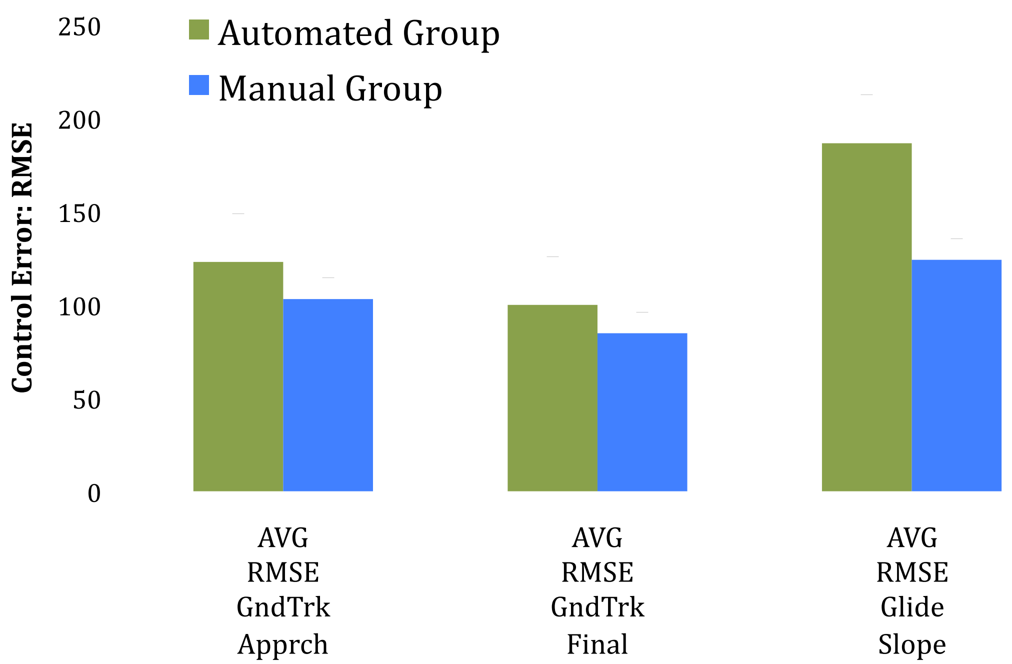


Figure 6
Root Mean Square Error for flightpath of UAV simulator as a function of prior type of training (manual control versus automation supported). Variability from the designated approach to the airstrip (GndTrk Apprch), course for final approach (GndTrk Final), and the angle of descent (Glide Slope) were recorded. These data show greater error on the glide slope for individuals trained previously with automation supporting aspects of their performance.



CONCLUSIONS

Across a series of experiments, our data consistently show no benefits for learning from the presence of automation during training, and frequent situations in which the presence of automation can impair learning. While automation does assist novice operators early in training, it apparently often does so at a cost to the degree of learning that occurs during training.

Moreover, the presence of an aptitude-automation interaction, shown in this project for the first time, suggests the effects of automation on training are greater for lower aptitude individuals. While supplying greater support to such users, the presence of automation may be masking differences between individuals and at the same time impairing their ability to acquire fundamental knowledge about the operation of the system. These are clearly matters of potential practical importance within numerous training situations.

Within the context of the current MURI effort specificity of training principle – in which learning and transfer are reduced when conditions within training differ from those encountered within a test (e.g., Healy & Bourne, 1995; Healy et al., 1993). Such a principle would suggest that changes to the nature of the task induced through the use of automation provide sufficient differences to impact performance when automation is withdrawn.

One possibility is that deficits in learning from automation are a direct product of increased reliance on automation during training. Researchers in the past (Bainbridge 1983; Endsley & Kiris, 1995; Moray, 1986) have suggested automation can impair acquisition and maintenance of operators' skill and the development of accurate mental models of the controlled system. One of the main consequences of reduced direct contact with the system is what has been termed the "out-of-the-loop performance problem" (Endsley & Kiris, 1995). This effect has been previously documented in other settings (e.g., Billings, 1991; Moray, 1986; Wiener & Curry, 1980). After prolonged interaction with automation, operators have diminished ability to detect system failures and subsequently take over manual control.

One simple solution to reducing reliance on automation, gradually withdrawing such support, proved ineffective in our research. While there may be other approaches that serve to inoculate individuals against over-reliance on automation, these remain topics for future research.

Our findings might be taken to suggest that individuals are best trained without automation. However, given the interaction of automation with aptitude, we suggest a future course in which the use and level of automation in training any individual is matched to the nature of training and the type of task. For example, given the reduced impact of automation on high aptitude individuals, there may be advantages to maintaining the availability of automation within some training contexts. For highly complex systems, use of automation may be a fundamental skill to be acquired, or errors that might otherwise occur as part of learning (and may be in some sense beneficial to learning) may have catastrophic consequences. Within systems where automation is available to support lower aptitude individuals, it may be that systems need to be designed with the intent of maintaining automatic support even as individuals apparently improve in their performance.

Overall we offer this very general, one line summary of our findings: Automation may seem like it is helping you, especially if you need more help, but it is not helping you learn.

Model evaluation: ACT-R, IMPRINT, and Matlab: Relative performances and new opportunities

Bengt Fornberg[1], William D. Raymond[2], Carolyn J. Buck-Gengler[2], Alice F. Healy[2], and Lyle E. Bourne, Jr.[2]

[1]Department of Applied Mathematics, University of Colorado, Boulder
[2]Department of Psychology and Neuroscience, University of Colorado, Boulder

Model evaluation: ACT-R, IMPRINT, and Matlab: Relative performances and new opportunities

Bengt Fornberg[1], William D. Raymond[2], Carolyn J. Buck-Gengler[2], Alice F. Healy[2], and Lyle E. Bourne, Jr.[2]
[1]Department of Applied Mathematics, University of Colorado, Boulder
[2]Department of Psychology and Neuroscience, University of Colorado, Boulder

1. Introduction and short summary of results.

Model comparison and model evaluation were issues of active interst prior to 2005 and the initiation of the present MURI (see, e.g., Gluck & Pew, 2001; Pitt & Myung, 2002; Young, 2003), but progress was at best limited. Thus, one goal of the present MURI was to attempt to develop the techniques and procedures needed for model comparison and evaluation and to demonstrate their utility with a set of new models designed specifically to predict training effects. This goal has, at least in part, been achieved, and the present report summarizes the significant advances that have resulted from the MURI effort.

This report summarizes the model evaluation effort within the present MURI. Our focus has been on evaluation of models of two tasks, a simple keystroke data entry task and a more complex visual search task (RADAR). Across both tasks, three different computational systems have been applied, ACT-R, IMPRINT, and Matlab, and the models in those systems compared.

Model evaluation has included model fits as well as timing comparisons of the three models. Model fits have been measured by comparing model simulations of the tasks against experimental data. Timing comparisons of the model simulations have been carried out on comparable computer systems (typically standard desktop and notebook PCs, using a single processor with clock speeds around 2-3 GHz).

The most striking outcome of the present model evaluation effort was the very large speed gains that proved possible when using the Matlab environment to model the tasks. Speed increases at factors around 10,000 were accomplished using Matlab. Equivalent, or perhaps larger, gains are likely if other scientific/engineering computer environments are employed, such as Fortran or C++. As a result of the speed increases using Matlab, the present model evaluation effort was extended to also explore the new opportunities increased model execution speed afforded in terms of: (1) performing "automated" parameter optimization; and (2) using radial basis functions (RBFs) to build computationally even faster approximations of the previously mentioned Matlab models' parameter spaces.

The model evaluation effort throughout the first half of the MURI has been described earlier (Fornberg, Raymond, & Best, 2007). The present report summarizes the complete model evaluation effort, although with a strong focus on the new opportunities that the previous work has opened up. It should be noted that these opportunities have a very direct impact on one of the main original questions, namely, the *scalability* of the present kinds of models. The major opportunities shown here to be available algorithmically (e.g., with a present Matlab RBF model (see Section 6) running some 5,000,000 times faster than a direct simulation in IMPRINT), together with additional factors in the hundreds or more readily available through the use of parallel computing (see Section 7), suggest that the

issue of *scalability* has been found to be not nearly the concern that it was perceived to be at the beginning of our MURI effort. Our main conclusion, however, is that different computational systems have different strengths and weaknesses, and each system should be left to handle what it excels in. In particular, scientific programming environments (such as Matlab, Fortran, or C++) can handle equation-based tasks with vastly greater efficiency than systems with other primary goals. On the other hand, IMPRINT is more appropriate for other military applications and ACT-R for modeling cognitive processes. With the ability of most systems to communicate data and
interchange computational requests, hybrid solutions will be needed to achieve the best results.

2. Modeling tasks.

Keystroke data entry and RADAR were chosen as the tasks for model evaluation for several reasons. Both tasks had been explored in multiple empirical studies. The many data entry studies, in particular, provided a rich source of data for modeling. The two tasks also differ in complexity; data entry is a cognitively simple task, and RADAR is a more cognitively challenging, complex task. In addition, these two tasks had both been modeled in ACT-R and IMPRINT. The models were developed by experts in the two platforms, which eliminated any potential concerns about non-optimal code implementations that might cause inappropriate biases in the comparisons. The ACT-R models were developed by the Carnegie Mellon team, with Brad Best as the primary programmer. For IMPRINT, the primary programmers were Carolyn Buck-Gengler and Bill Raymond at the University of Colorado, Boulder. The (equation-based) IMPRINT models were subsequently converted to Matlab by Bengt Fornberg and Bill Raymond. Both of the models focused on cognitive phenomena of the data entry and RADAR tasks, which are directly relevant to the effects of training on performance. The models were developed not only to give us descriptive and predictive capabilities, but also to deepen our understanding of the genuine nature of the processes that are modeled.

2.1. The keystroke data entry task.

The first model comparison problem involved the keystroke data entry task in experiments described in Healy, Kole, Buck-Gengler, and Bourne (2004). ACT-R and IMPRINT models of this task are described in Gonzalez, Fu, Healy, Kole, and Bourne (2006) and in Buck-Gengler, Raymond, Healy, and Bourne (2007). The experiments required subjects to type 4-digit numbers that were displayed to them on a computer screen. They were instructed to type the numbers as quickly and as accurately as possible. Numbers were presented one at a time and were typed on the keypad to the right of the keyboard. Subjects did not see their typed numbers, and they terminated each trial by pressing the “Enter” key. The stimuli consisted of 10 blocks of 64 numbers each, which were divided by a short break into 2 session halves of 5 blocks each. In both experiments there were 32 subjects. In Experiment 1, a set of 64 numbers were repeated in each of the 5 blocks of the first half in different random orders, and a second set of 64 numbers were repeated in different random orders in each of the 5 blocks of the second half. All subjects typed using their left (non-dominant) hand. In Experiment 2, all numbers were unique, and the hand used for typing

(Left, Right) was crossed with session half to create 4 conditions of hand use (LL, LR, RL, and RR).

2.2. The RADAR task.

This second model comparison problem, and its ACT-R and IMPRINT implementations, are described in Best, Gonzalez, Young, Healy, and Bourne (2007), in Young, Gonzalez, Dutt, and Bourne (in press), and in Buck-Gengler, Raymond, Healy, and Bourne (2010). We will not repeat the description of the task in any detail here (or address its conceptual significance), apart from noting that it combines mapping type (targets and foils from same or different character sets), load level (number of items in target set; number of items to look at to see if a target), and tone counting (concurrent secondary task using auditory modality). In the experiment, 12 subjects performed two *sessions* of eight *blocks*, each with 20 *shifts* of 7 *frames*.

3. Modeling principles and platforms .

3.1. Modeling principles.

As noted in Fornberg et al. (2007), there are two fundamentally different modeling methodologies that one can apply to a modeling problem, which we in short (and grossly oversimplified) denote by *first principles* and *brute force data fitting*. A first principles approach begins with a theory of the principles that underlie the phenomena to be modeled, in this instance human performance on a task. This approach was pursued in developing both the ACT-R and the IMPRINT/Matlab models. A first principles approach is highly desirable when it works well, that is, when the principles are known, which is the case for the cognitive tasks presently being modeled. However, its successful application depends on the nontrivial task of developing a theory for a situation of intrinsically very high complexity. Brute force data fitting is closely related to the process of *data mining*. This approach is a fairly new and very active general research area. The strength of the data fitting approach is its ability to bring out entirely unanticipated, but nevertheless significant, relations in the data. Such relations frequently lie deeply hidden in most large data sets, and virtually always escape attention when using conventional visualization or similar inspection methods. Novel approaches to data fitting include the use of neural networks and radial basis functions (RBFs). Once such a model has been created, another advantage to it is that it can be evaluated extremely rapidly.

In the last year, we have followed up on brute force data fitting by creating an RBF approximation of the Matlab model’s parameter space for the keystroke data entry task. We carried out this exercise by evaluating the first principles model quite a large number of times and then used the resulting data to obtain the second model, which we therefore can describe as a “model of a model.” As will be described later (§6.1), this procedure allows the elimination of stochastic noise and, more importantly, allows for very much faster model evaluations. As one application, we describe in Section 6.2 how this exercise can be used for interactive visualization of functions that depend on many parameters.

3.2. The modeling platforms.

Although the details of the ACT-R and the IMPRINT/Matlab first principles models differ fundamentally, implementations in all the three programming environments share a number of underlying general principles, including the facts that (1) the tasks are decomposed into simple conceptual components; (2) the components are combined to create a simulation with a (relatively) user-friendly interface; and (3) the generated data simulate variable human behavior on the tasks. However, the modeling platforms differ significantly in several respects. The focus of intended use is different for the platforms: ACT-R was designed for cognitive modeling; IMPRINT was designed for assessing human performance in military tasks; and Matlab was designed for science and engineering applications. The platforms also differ in raw computational speed, with Matlab faster than the other two platforms. In addition because Matlab was intended for general engineering and scientific use, it has available within it a number of tools for carrying out parameter optimization, graphics, and interfacing to parallel computing hardware, which the other platforms lack. On the other hand, because of their intended uses, both the ACT-R and the IMPRINT platforms have large amounts of human performance-specific information built in, whereas Matlab does not (although appropriate libraries could be added). This difference will inevitably make the former two platforms slower on some simple tasks, but gradually more powerful as this type of information is increasingly called for in more complex tasks or scenarios. Moreover, the specific embedded information differs in ACT-R and IMPRINT: ACT-R embodies a theory of general cognitive mechanisms; IMPRINT can call on information regarding the skills and abilities of army personnel engaged in military tasks.

The next three subsections give brief comments on the three modeling platforms, following the description given earlier in Fornberg et al. (2007). The Appendix in this earlier work contained illustrations of code for the three systems.

3.2.1. The ACT-R modeling platform.

ACT-R (Anderson, Bothell, Byrne, Douglass, Lebiere, & Qin, 2004; Anderson & Lebiere, 1998) is a unified theory of cognition developed through over 30 years of cumulative improvement. At a fine-grained scale it has accounted for hundreds of phenomena from the cognitive psychology and human factors literature. The version employed here, ACT-R 6.0, is a modular architecture composed of interacting modules for declarative memory, perceptual systems such as vision and audition modules, and motor systems such as a manual module, all synchronized through a central production system.

ACT-R is a hybrid system combining a tractable symbolic level, implemented as a production system that enables the specification of complex cognitive functions, with a subsymbolic level that tunes itself to the statistical structure of the environment. The combination of these aspects provides both the broad structure of cognitive processes and the graded characteristics of cognition such as adaptivity, robustness, and stochasticity.

The central part of the architecture is the production module. A production can match the contents of any combination of buffers. Buffers include the goal buffer, which holds the current context and intentions, the retrieval buffer, which holds the most recent chunk retrieved from declarative memory, visual and auditory buffers, which hold the

current sensory information, and the manual buffer, which holds the current state of the motor module. During the matching phase, production rules whose conditions match to the current state of various information buffers (goal, memory retrieval, perceptual, etc.) qualify to enter the conflict set. Because ACT-R specifies that only one production can fire at a time, the rule with the highest expected utility from among those that match is selected as the one to fire. Utility is graded both by the expected value of information, driven by activation, and the quality or exactness of the match itself.

The general structure of the ACT-R models used in the following data entry experiments includes two main steps: 1) noticing and encoding of the stimulus from the computer screen, and 2) entry of the encoded stimulus using the keypad. The first step further unpacks to include reading of individual numbers, whereas the second step includes preparing the proper motor program to press the desired keys. These steps say little about whether numbers are encoded more than one at a time, and whether any key presses occur before all of the numbers are encoded. As is described below, human participants actually use multiple strategies to approach even this simple task, and tend to vary between individuals in a preference to either encode all four digits before entering any, or to encode a pair of digits at a time, entering a pair after it is encoded. Thus, the model was constructed to support both of these strategies. Again, though the task is quite simple, it still requires maintenance of encoded stimuli in working memory, potentially decomposing a task into subgoals (working on entering one pair at a time), and the interaction with skilled actions (keyboard entry), which is simulated through the application of individual ACT-R productions (e.g., typing the “9” key on the keypad).

3.2.2. The IMPRINT modeling platform.

The versions of IMPRINT used for MURI modeling, IMPRINT 7 and IMPRINT Pro, are primarily used to create simulations of military personnel and equipment engaged in military tasks. The simulations can be used to evaluate planning efficiency, given constraints on time, accuracy, and equipment functionality, as well as human skills, abilities, and capacities. Simulations can also take into account variables in the external environment that may affect personnel or equipment. IMPRINT was not specifically designed for modeling cognitive tasks; however, the current modeling effort shows that cognitive models can be implemented on the IMPRINT platform.

The IMPRINT model of the keystroke data entry task was based on a cognitive model of the task that involves three subprocesses:

(1) Read and represent a number: Read each number and create an ordered mental representation of the digits, one digit at a time;
(2) Create motor plan: Access each of the represented digits in sequence to create a motor plan for typing it; and
(3) Execute motor plan: Utilize the motor plan to type each digit, followed by the enter key.

The subprocesses were assumed to occur sequentially for each number. However, accommodation was made in the simulation for a phenomenon observed in the experimental data in which some subjects tended to group, or *chunk*, the first two digits of

a number and the last two digits of a number, as evidenced in these subjects by longer response times for the third keystroke than for the second and fourth. The chunking phenomenon presumably entails some additional cognitive processing between the two chunks, which was simulated in the model.

The IMPRINT model consisted of a main network and a goal network. In the main network, parameters can be set to duplicate the conditions of Experiment 1 of Healy et al. (2004) (all left hand typing and number repetition in each half) or of Experiment 2 (typing hand crossed with session half and no repeated numbers). The goal network was called repeatedly until the stimuli were exhausted. Each run of the model represented the output from one statistical subject.

A number of human performance parameters in the model were assigned values stochastically to simulate human variability of performance. Values for stochastic variables were taken from a variety of probability distributions (viz., normal, uniform, and gamma), which were chosen, together with their parameters, to capture distributions observed in the experimental data. Other model parameters were predetermined through data inspection and do not vary in the model.

3.2.3. The Matlab modeling platform.

Matlab evolved from FORTRAN in the late 1970's, and has since become one of the most widely used programming environments in science and engineering. The language is technically an interpreted one, but its statements are in effect compiled on their first execution, and then reused in this latter form. The language is built around matrix (or vector) operations and, when used in such way (as opposed to in scalar form with many nested loops and conditional statements), its speeds normally come quite close to what the computer hardware is theoretically capable of.

The hardware of modern PCs often allows many computational threads to execute simultaneously. Not only are computers typically equipped with one or several dual-core (or multiple-core) processors, each of these cores may furthermore be hyper-threaded (doubling again the number of independent simultaneous threads). The resulting opportunities of parallel processing are automatically utilized in Matlab's matrix operations, with no special user attention needed. Matlab's parallel computing toolbox can be used to utilize parallel processing also for other types of operations with (in most cases) only a few lines of extra programming. This feature is discussed further in Section 7.

In the present project, the Matlab model was a direct translation of the algorithms used in the IMPRINT code. We have found several advantages in porting the numerical parts of IMPRINT codes to Matlab. Importantly, Matlab has very much higher execution speeds than IMPRINT. Matlab code is also short and easy to write. It can also be comprehensively viewed as a single program, unlike IMPRINT code. As mentioned, Matlab also has available within it some powerful tools for graphics, optimization, debugging, and profiling (i.e., code timing). Modeling in IMPRINT also provided us the ability to compare environments specific to modeling human cognition and performance (ACT-R and IMPRINT) against one with no such specialization, but instead focused on high speed computing.

We want to stress again that the choice of Matlab (as opposed to, say, FORTRAN, C++, or Python) was made for obtaining outside benchmark assessments on the evaluation

speeds of ACT-R and IMPRINT in the most flexible and convenient way possible, and not because we expect this particular language (Matlab) to be adopted on a large scale by the military for cognitive modeling. Matlab simply allowed our focus to be strongly concentrated on obtaining timing and scalability comparisons with the least possible attention diverted to implementation technicalities.

4. Model comparisons.

Model comparisons were performed in two ways. First, for a model to be useful, it must be fast to run. Thus, we collected performance timing information for execution of each model on each test problem. Second, the model should be capable of accurately simulating the empirically derived human data. To measure the models' reliability in this regard, we obtained correlations among the model outputs and the experimental data, for each experiment. In addition, Root Mean Square Errors (RMSEs) were calculated between models and the experimental data.

Given that the relative timing was similar for the two (quite different) modeling tasks (in terms of relative performance between the three computational systems), it suffices to focus here on one of the tasks, keystroke data entry.

At the time of working with this first test problem, we were concerned that this problem might be misleadingly favorable to Matlab, because its logical structure was such that all the arithmetic operations of the IMPRINT model could be recast into Matlab's matrix syntax (in which case Matlab is particularly computationally efficient). This opportunity of using matrix syntax was not available for the RADAR task, but we nevertheless found equivalent relative speed differences, assuring us that the general observations we are making are not due to any such special circumstances.

Accuracy comparisons between the ACT-R and IMPRINT models have been reported separately (Fornberg et al., 2007) and will not be repeated here, apart from noting that generally, there were no significant differences. The Matlab code for the keystroke task was a quite direct translations of the IMPRINT one, whereas the Matlab code for the RADAR task was structured differently as nested loops, rather than as many separate modules interacting with each other. In both cases, the mathematical algorithms and parameter choices were identical, and hence, there were again no differences in modeling accuracy between the IMPRINT and Matlab versions. Since all models of the present kind rely heavily on random samplings, no two runs will give identical outputs. The differences between IMPRINT and Matlab runs were no larger than between two different IMPRINT runs or between two different Matlab runs.

4.1. ACT-R performance timing on the keystroke task.

The ACT-R model was run on a Dell laptop running Windows XP with a 2.0 GHz mobile Intel CPU and 1 GB RAM. ACT-R requires a Lisp environment as well; the current model runs used Allegro Common Lisp version 6.1 and ACT-R version 6.0. It is worth noting in a section on timing that Lisp is an interpreted programming language, and, as a result, optimization techniques (which have not been used here) can produce significant speedup through allowing code to be partially compiled. It is also worth noting that, in addition to the small amount of data that are collected and collated for the individual model runs, the

ACT-R system processes and records a large amount of information that was not used in the current study, but is nonetheless readily accessible (e.g., activations of every chunk, previous instantiations of productions, a record of every goal the system attempts to achieve, etc.). Stated differently, the performance data produced by the model are derived from its behavior rather than produced as a primary product.

The following is a summary of the time requirements to run a batch of 32 simulated participants on the system described at the start of this section (as printed out by the ACT-R system)

```
cpu time (non-gc) 605,045 msec (00:10:05.045) user, 392 msec system
cpu time (gc)     251,940 msec (00:04:11.940) user,  77 msec system
cpu time (total)  856,985 msec (00:14:16.985) user, 469 msec system
real time  859,952 msec (00:14:19.952)
```

The time is broken up into 'garbage collection' (gc: a Lisp system activity) and actual program execution (non-gc), and then summed into a total time for the processing. This sum places an upper bound of approximately 30 seconds on the processing time required to simulate an individual participant (as well as accomplish the file I/O to write out detailed data files containing individual trial results for each participant and their summaries and handle the memory management necessary for the Lisp interpreter).

4.2. IMPRINT performance timing on the keystroke task.

The IMPRINT model writes output data to a Microsoft Excel spreadsheet. For the timing comparisons that are reported here, we have not included that overhead, but only the time needed for producing the means for each of the 10 blocks, when averaged over all statistical subjects (in each condition) and over all non-error items, for both experiments.

The IMPRINT code (running version r7.30 on a Dell computer under Microsoft Windows XP Professional with a 2.8 GHz processor and 2 GB memory) required 24 minutes for each experiment. This total amounts to approximately 45 seconds for each of the 32 subjects. Writing all the generated data to an Excel spreadsheet file takes an additional 10 minutes. IMPRINT does not include any profiler option that details how much time each line of code takes. The times quoted are “wall clock times”.

4.3. Matlab performance timing on the keystroke task.

The code for the Matlab implementation of the data entry required no more than about 70 lines of code (not counting comment lines). Execution of the code with parameters set for simulation of Experiment 1 (all 32 subjects) took approximately 0.085 seconds on a Dell GX-270 PC single processor operating at 3.2 GHz, with 2 GB RAM, running under Windows XP. The time thus becomes about 0.0027 seconds (2.7 ms) per subject. The *profiler* of Matlab allows for very convenient and detailed timing of codes, showing in particular how much time is spent on each line of code. A condensed output (listing only the lines taking the most time) is seen below:

Lines where the most time was spent

Line Number	Code	Calls	Total Time	% Time	Time Plot
105	g_dist = (0.5*sum(randn(nr_c,8...	320	0.025 s	29.1%	
116	total_times = ...	320	0.017 s	20.3%	
113	CC_PH_Actual = (Counts(:,1)-Co...	320	0.010 s	11.7%	
73	Counts = cumsum([[ItemCt,Corre...	320	0.006 s	7.1%	
71	ab = 3-s1-s2;	320	0.003 s	3.5%	
All other lines			0.024 s	28.3%	
Totals			0.085 s	100%	

The timing for Experiment 2 was equivalent, resulting in a typical computer time of 0.17 seconds for running both cases.

The ratio between the Matlab and IMPRINT execution times is thus similar to 6/100,000. Because the speeds of the computers are roughly similar, this ratio suggests that porting specific subtasks to Matlab can offer gains that are much larger than, say, porting from a PC to a giant supercomputer system. We wish to stress that using different computer languages/systems for different tasks within a single project is a much more appropriate approach for large tasks than to rely on a single language only. Most languages/systems have interface options to run sub-tasks in other languages. For example, IMPRINT and Matlab have built-in facilities to execute modules in C++ or Fortran.

4.4. IMPRINT and Matlab performance timing on the RADAR task.

With regard to the conversion of IMPRINT to Matlab, the RADAR task differed from the data entry task in two primary ways: (2) stochastic features enter in the RADAR task in such a way that Matlab's array processing features are no longer practical to apply; and (2) the general programming style of Matlab (shared with Fortran, C/C++) allows for particularly simple and effective code structure with nested loops instead of many interacting modules. The two issues led to roughly offsetting advantages and disadvantages. The Matlab model turned again out to be about 10,000 times faster than IMPRINT on equivalent hardware. The Matlab code was again extraordinarily compact and readable - this time about 100 lines only of executable code (plus about 30 lines for entering parameters). A typical output of a Matlab RADAR simulation run (with the timing displayed) is given in the context of parameter optimization in Section 5.2.

5. Parameter optimization.

The model evaluation requirement in the original proposal was mainly limited to accuracy and performance comparisons and to assessment of scalability for future larger modeling tasks. However, the extreme speed advantages of the Matlab implementations pointed immediately to several opportunities that were not originally anticipated. The first of these concerns parameter optimization.

There are several approaches available for parameter optimization, some of which are yet to reach their full potential in cognitive modeling environments. Although finding the *global* optimum of a function of 1 or 2 variables usually can be handled efficiently, and finding *local* optima of functions of many variables is also relatively straightforward (meaning that effective algorithms and software is available in optimization "toolboxes"), the issue of finding global optima of functions of many variables is a daunting one. In Gluck, Scheutz, Gunzelmann, Harris, & Kershner (2007), a calculation based on ACT-R, exploring a 4-variable parameter space by means of 21, 26, 105 and 31 increments in respective parameters, is reported to have consumed 96,000 processor hours on a cluster at Wright-Patterson's High Performance Computing Center. For each additional variable sampled in a similar manner, times would be expected to rise by another factor of around 20. Clearly, it is critically needed both to use faster general optimization algorithms and to increase the computational speed as far as possible.

In a famous review in 2000 of the 10 most influential algorithms developed during the 20th century (Cipra, 2000; Dongarra & Sullivan, 2000), *simulated annealing* appeared in first place. *Genetic algorithms* is a second approach, whose full impact is yet to be fully experienced. In many cases, optimization even over dozens or tens of dozens of variables can be entirely feasible. Both simulated annealing and genetic algorithms are available within the Matlab environment, and adding either of these optimizers 'on top of' an existing model requires less than 10 extra lines of code.

5.1. Two illustrations of displaying multivariate functions.

Figure 1 illustrates in generic form, with a 3-D function not related to the concept of cognitive modeling, an intrinsic problem with displaying functions with more than two independent variables.

In Figure 2 we consider the data entry test problem, with the following five key parameters in Table 1 as independent variables and display the RMSE (root mean square error) between the IMPRINT/Matlab model as the dependent variable (objective function value). In a 5-D space, we can make 10 2-D slices if we lock in three variables at a time at their "hand derived values" and let the two remaining parameters vary over their respective "reasonable ranges." Figure 2 displays these 10 slices in two different ways: as contour plots and as surface plots. To a much larger extent than for the 3-D function in Figure 1, these slices give an extremely incomplete picture of the full 5-D functional dependence of the RMSE. The optimization task we are addressing is to carry out a complete 5-D space minimization of the RMSE (i.e., not limited to these "slices").

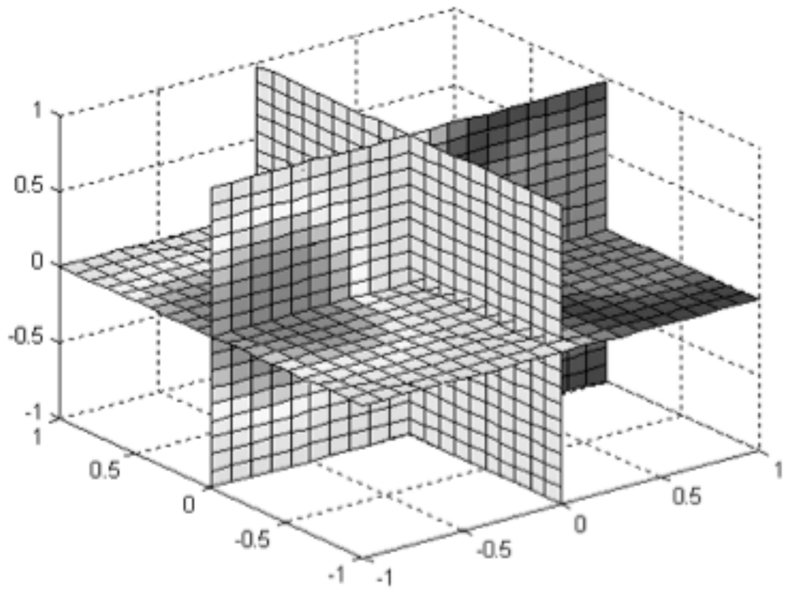


Figure 1. *Schematic illustration of a function of three independent variables. In contrast to a function of two variables (Figure 6), we cannot display a function of three variables on a flat paper or computer screen throughout its full domain, but need to limit ourselves to showing only "slices" of it, omitting potentially critically important areas.*

Table 1. *Five key parameters of the data entry model selected for optimization, along with reasonable ranges for each parameter and the original IMPRINT model's hand-derived values.*

Parameters selected:	Reasonable ranges	Hand derived values
Cognitive learning (cognitive)	[-0.20, 0.00]	-0.045
Motoric learning (physical)	[-0.10, 0.00]	-0.015
Repetition priming (rep learn)	[0.00, 0.30]	0.050
Left-hand penalty (left penalty)	[1.00, 1.50]	1.125
Cognitive slowdown (fatigue)	[0.00, 0.10]	0.0125

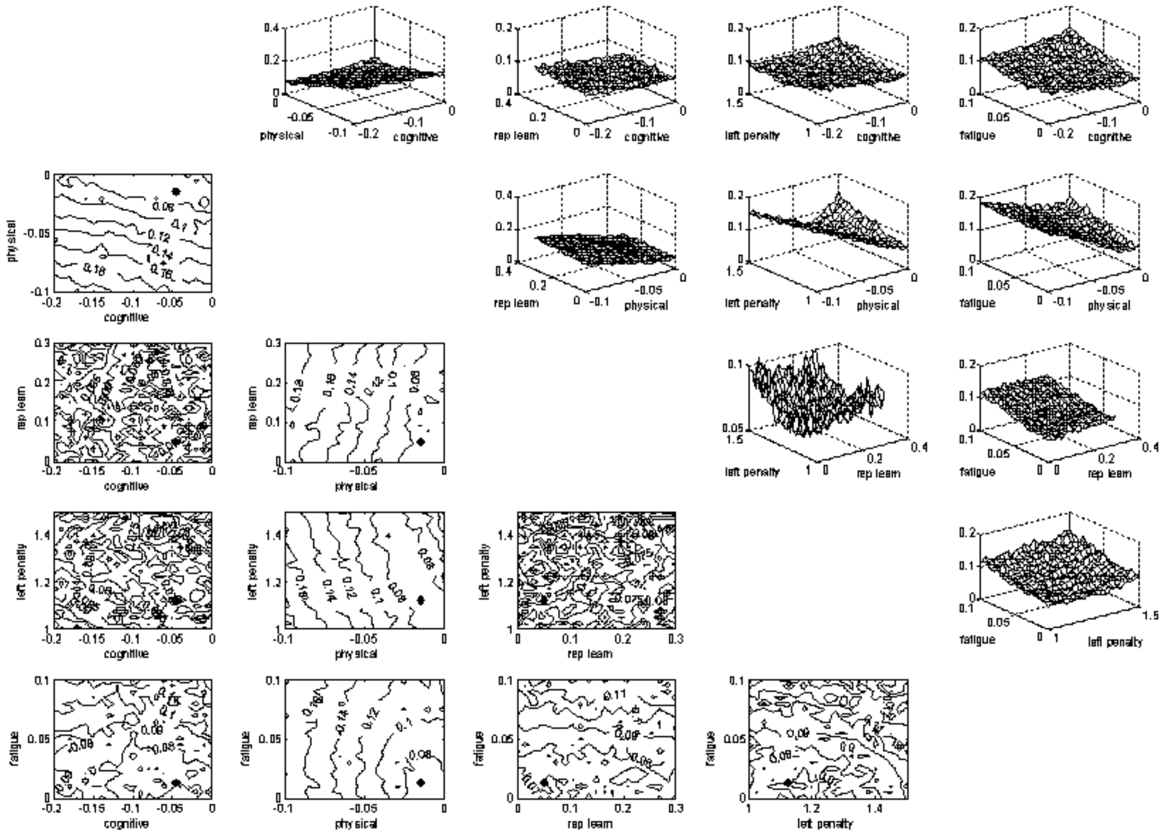


Figure 2. *Displays of RMSE errors for the Matlab keystroke model when any pair of 2 out of the 5 parameters was left to freely vary over their 'reasonable range', while the remaining 3 parameters were held at their 'assumed best' positions. The same data are displayed in the top right and the bottom left subplots; as surface plots and as contour plots, respectively. The solid dots in the latter figures mark the 'assumed best' values. The fact that these dots are seen to be located at low spots of the different functions indicates that the (time consuming) manual parameter determination was successful. However, the 10 different parameter space slices shown here explore only a minute fraction of the full 5-D parameter space, leaving completely open the possibilities of much better parameter combinations in other parts of that space.*
`

5.2. Parameter optimization using genetic algorithms (GA) and simulated annealing (SA) on the data entry task.

A large number of optional *Toolboxes* are available with Matlab, including one that provides both Genetic Algorithms (GA) and Simulated Annealing (SA) capabilities. As noted above, these are two very successful strategies for searching through high-dimensional parameter spaces for locating global optima more effectively than an exhaustive parameter space search. Both search methodologies borrow their key ideas from processes in nature: biological evolution (for GA), and crystal formation through slow cooling (for SA). For the results shown in Figure 3, GA and SA optimizations were each run 20 times. These runs executed both Experiments 1 and 2. A GA optimization consisted of letting a population of size 30 evolve through 60 generations, for a total of 1800 evaluations of the RMSE objective function. The typical time for each GA optimization

was about 5 minutes. The SA optimizations were stopped after about equally many objective function evaluations, thus again taking around 5 minutes each full run.

The 20 GA and 20 SA runs give results comparable to what an exhaustive search would have provided (but in a fraction of the approximately 10 days the latter would have required).

For each of the 5 selected variables, we see displayed two horizontal lines with short vertical lines between them. The extent of each of the horizontal lines corresponds to the "reasonable range" for the respective variable, as shown at the left edge. Along the top line for each variable, we see the outcome for that variable of 20 separate global GA optimizations, and along the bottom line, the same for 20 SA optimizations. Due to the very flat character of the function that is optimized, together with its large noise level, the results are very satisfactory in showing

- The manually found values indeed are consistent with global optimization results.
- Thorough manual optimization (feasible here, but not always practical) can be confirmed (or replaced) by merely minutes of computing using a global optimizer.
- The variation between different optimization runs can provide good information about different model parameters' uncertainty ranges.
- The presence of even large amounts of statistical noise in a model does not cause major difficulties for fully automated parameter determination with either GA or SA.

Since scaling issues form a critical aspect of the present model evaluation task, we can note that having 10 parameters instead of 5 in the optimization would only increase the GA or SA times by a factor in the 20-100 range, whereas the cost for an exhaustive search would increase times by a further factor of 21^5, that is, to completely unrealistic computer times of the order of 500,000 years.

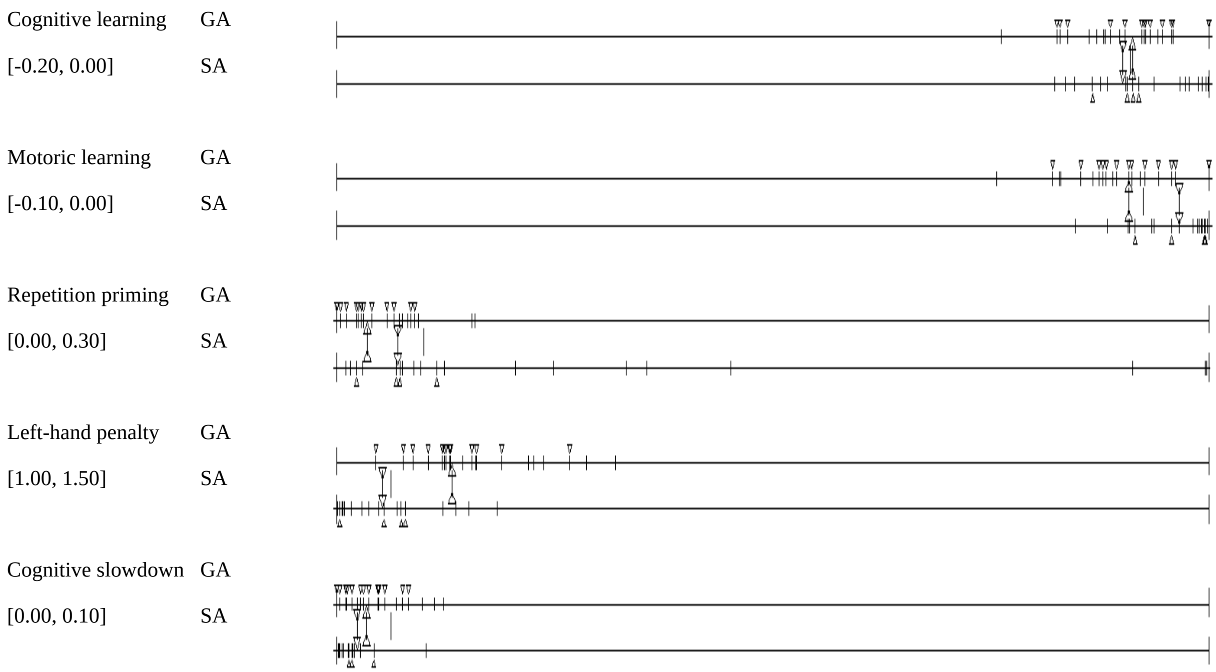


Figure 3. *Outcomes of 20 GA and 20 SA optimizations of 5 model parameters. The horizontal lines represent the search ranges. Short vertical lines indicate the outcomes of individual optimizations, with small triangles pointing at those yielding particularly low values of RMSE (<.06; averaged over the 2 experiments). The vertical line segments with triangle pointers at each end show the average GA and SA results based on the low-RMSE results. We can see that these optimal values mostly are in fine agreement with the hand-derived parameter values (vertical lines with no end markers).*

5.3. Parameter optimization by genetic algorithms (GA) on the RADAR task.

The lines below show a typical output of one single run of the Matlab RADAR modeling code based on the IMPRINT model (here executed on a Dell D430 notebook computer, with a 1.33 GHz processor):

```
Model_RT =
  0.6130  0.9242  0.6261  1.1041  1.1854  0.6123  0.9189  0.6116
  0.6310  0.9213  0.6348  1.1569  1.1446  0.5989  0.9293  0.6036

Experiment_RT =
  0.6169  0.9293  0.6188  1.1830  1.1181  0.6187  0.9775  0.6622
  0.6176  0.9039  0.6318  1.1249  1.0735  0.6271  0.8511  0.6164

RMSE_RT =
  0.0437

Model_hits =
  0.9944  0.9778  0.9244  0.7379  0.8290  0.9313  0.9885  0.9944
```

```
  0.9889  0.9889  0.8762  0.7873  0.7828  0.8873  0.9778  0.9889

Experiment_hits =
  0.9667  0.9663  0.9761  0.7515  0.7762  1.0000  0.9885  0.9944
  0.9944  0.9833  0.9833  0.7901  0.7955  0.9829  0.9889  1.0000

RMSE_hits =
  0.0450

Model_FA =
  0.1750  0.1226  0.1470  0.1780  0.1364  0.0972  0.1009  0.0544
  0.1198  0.0286  0.1428  0.0952  0.1114   .0966   .0591  0.0119

Experiment_FA =
  0.1663  0.0278  0.1210  0.2142  0.1012  0.1472  0.0821  0.0556
  0.1000  0.0544  0.1444  0.0694  0.1040  0.1000  0.0306  0.0472

RMSE_FA =
    0.0345
Elapsed time is 0.154991 seconds.
```

The details of what the numbers in the computer output above stand for is not the essential point here, but rather:

i. The only experimental data that is provided to compare the model against is given under the headings `Experiment_RT` , `Experiment_hits` , and `Experiment_FA` (scores for the quantities *response times* (for hits), proportion of *hits*, and proportion of *false alarms*, respectively, averaged over *subjects*, *shifts*, and *frames*). The two rows for each variable represent the two *sessions*, and the eight columns represent the eight *blocks*. In all, the supplied experimental data amounts only to 48 numbers.

ii. The RADAR model, for each run, produces a different set of matching 48 numbers, shown in the output above under the headings `Model_RT`, `Model_hits`, and `Model_FA`, respectively.

For each run of the model, we can trivially calculate the RMSE error in each of the three categories and then form an overall RMSE as the average of these three. In the single instance listed above, this gives

RMSE = (`RMSE_RT` + `RMSE_hits` + `RMSE_FA`) / 3 = 0.0410.

There is a quite high level of stochastic fluctuations in the model and, when averaging over 100 runs, the average RMSE turns out to become significantly larger: RMSE = 0.0434.

The RADAR model contains about 30 nontrivial parameters. Rather than attempting a global optimization simultaneously over all of these (if so, needlessly making a very challenging problem nearly impossible), we can select out groups of parameters that logically belong together, and for which the hand-derived values are particularly uncertain (or particularly interesting). As an example, we select here 16 parameters in four groups:

1. Four parameters for `DecisionTimeDist`
2. Four parameters for `ResponseProb`
3. Five parameters for `FABlockTypeRate`

4. Three parameters for `FALearningRate`

Continuing to omit technical details in this summary in order to convey the key concepts more clearly, we do not here describe just what these groups and their parameters represent from a motoric/cognitive perspective, but proceed instead with describing the outcomes of our GA optimizations in these four cases. In each case, we ran 20 GA optimizations with *population sizes* of 40 that evolved through between 10 and 30 *generations* (with less needed when the parameters were fewer). The result is illustrated in Figure 4. In each of the four subplots, we see one horizontal line for each parameter. To the left is given a 'reasonable range' and below each line a small vertical tick mark shows the hand-derived value. Above each line, we see similarly the outcomes of the 20 GA simulations. In most cases, the agreement is fully satisfactory, but we see also instances of exceptions (e.g. the third case in the top right subplot and the last two cases in the bottom right subplot). In some cases, the parameters turn out to be well determined by the GA data whereas, in other cases, the uncertainties are large.

Adjusting the model parameters to agree better with the GA results, for example replacing each value with the average for the GA runs, reduced the typical RMSE by about 15% (to around 0.0373 when averaged over 100 simulations, compared to the value 0.0434 quoted above). The level of reduction is not so much the issue as the fact that we, in a totally automated way and in spite of all the random fluctuations, can get information separately on a large number of parameters, although these contribute only in combined form towards the (measurable) model fit, as represented by the RMSE.

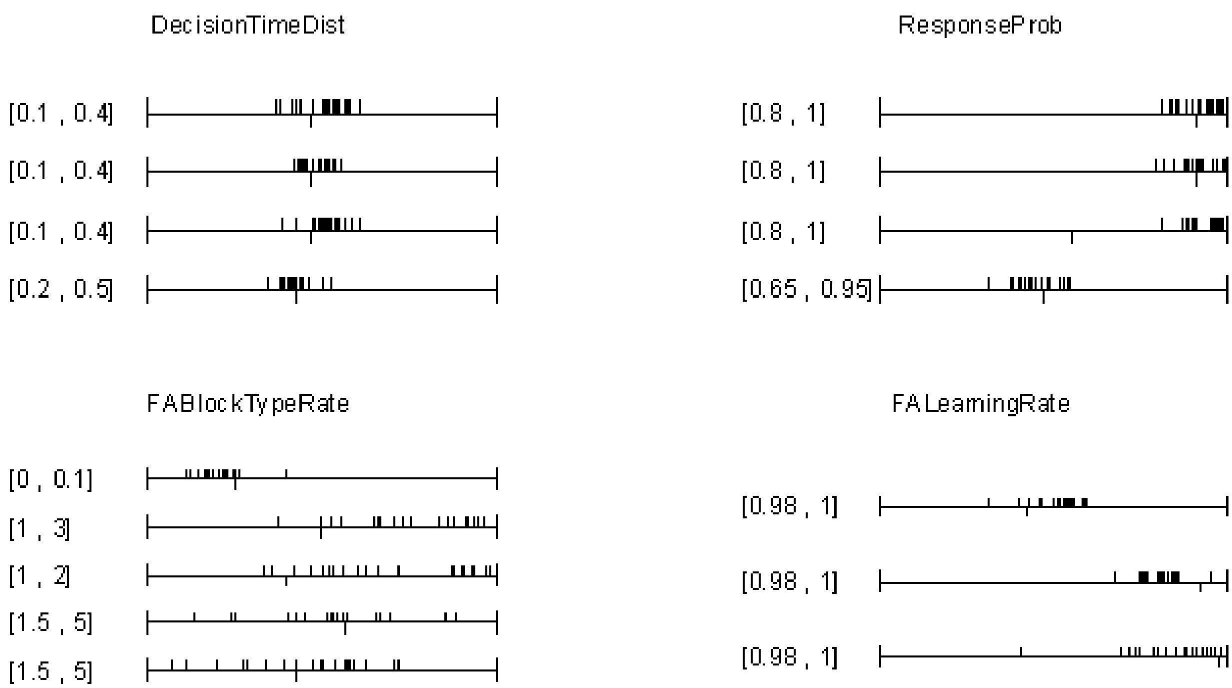


Figure 4. *Comparison between hand-derived and GA obtained values for 16 model parameters, distributed over four different parameter categories. For each parameter, a 'reasonable range' is displayed, with the hand-derived value marked below each line and the results of 20 GA optimizations marked above it.*

6. Radial Basis Function (RBF) models of Matlab models.

6.1. Concept and computational speed of RBF models

The methodology of Radial Basis Functions (RBFs) was first proposed about 40 years ago (Hardy, 1971) in a context related to that for which we will be using it here: multivariate interpolation. Its generality and power has only been fully recognized much more recently. It is nowadays successfully employed in numerous other areas of application, such as within neural networks, for the numerical solution of partial differential equations, and for graphical surface rendering. For a very good recent survey of the concept of RBFs, their mathematical background and approximation properties, as well as their efficient implementation in Matlab, see Fasshauer (2007). Note that multivariate interpolation allows the creation of very fast-to-evaluate RBF approximations of functions. In our case, we creatd an RBF model of the previous model's objective function used to optimize model parameters. We thus evaluated the previously developed Matlab model of they keystroke task at some (i.e., a few thousand) suitably chosen parameter locations and then, with about 20 lines of additional code, created an RBF model of the Matlab model's parameter space that reproduces the process of parameter space evaluation, but with stochastic noise suppressed to whatever degree we desire. Applied to the 5-parameter keystroke model, with the same 10 "slices" through its 5-parameter space illustrated in Figure 2, the RBF model will give the result shown in Figure 5. We immediately recognize the identical trends (which will hold throughout the full parameter space, and not just on the shown slices). The big advantages include:

1. Computational speed. Although each evaluation of the original Matlab model (both Experiment 1 and 2) required 0.17 seconds, the RBF model evaluates in 0.00030 seconds, i.e. over 500 times faster than the already fast original Matlab version.

2. Elimination of the stochastic noise. The GA (and SA) algorithms were highly effective for parameter optimization, even in the presence of the noise, so the gain achieved by working instead with smoother (and deterministic) functions proved to be minor. However, the speed gain of about 500 will make optimizations faster by about that same amount.

The computation behind Figure 2 required a total of 4410 evaluations of the Matlab model for each of "Experiment 1" and "Experiment 2." The total time for producing Figure 2 was 12.5 minutes (which would have been 147 days in IMPRINT). In contrast, the computing time for producing the data for Figure 5 (once the RBF model had been created) was 1.3 seconds.

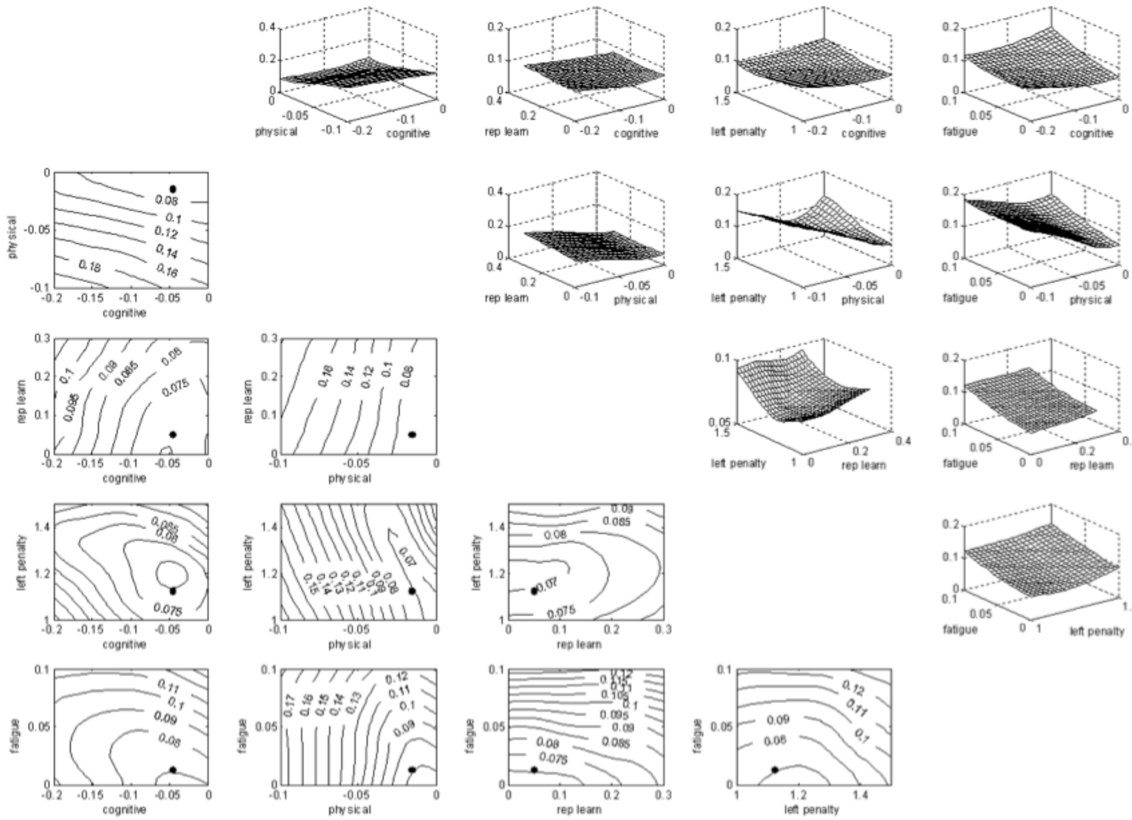


Figure 5. *The counterpart to Figure 2, with the difference that the original IMPRINT/Matlab First Principles model has been replaced by a RBF-type Brute force data fitting model. We visually recognize all the trends from the display in Figure 1, but the stochastic 'noise' has been successfully eliminated. Each functional evaluation in this model of a model is about 500 times faster than for the original model.*

6.2. Opportunity provided by RBFs for multivariate visualization.

The ability to evaluate an RBF model very rapidly opens up an opportunity - not yet utilized in the literature - to interactively move through different dimensions and thereby display multivariate functions without the customary limitation of 2-D paper or equally flat computer screens. The left part of Figure 5 displays a standard 2-D surface plot, conveying very clearly the character of a function of 2 variables. A dashed frame shows how one can “slice” out a 1-D function of x only (with its y-value fixed as a certain value y_0). This slice can be displayed as a curve shown to the right, together with a “slider” that can be moved by a mouse, causing the curve above it to dynamically update. By this method, we can visualize a 2-D function as a 1-D curve together with one slider. The opportunity that fast RBF models offer in this regard is that the function to be displayed can be in d-D. If we display a surface (2-D) and use d-2 sliders, moving these sliders allows immediate visual inspection of d-D functions. This is an entirely novel opportunity offered by the present d-D RBF models due to their very high computational speeds (effective up to d = 5 or 6, i.e. very well past the usual d = 2 or 3 limitation).

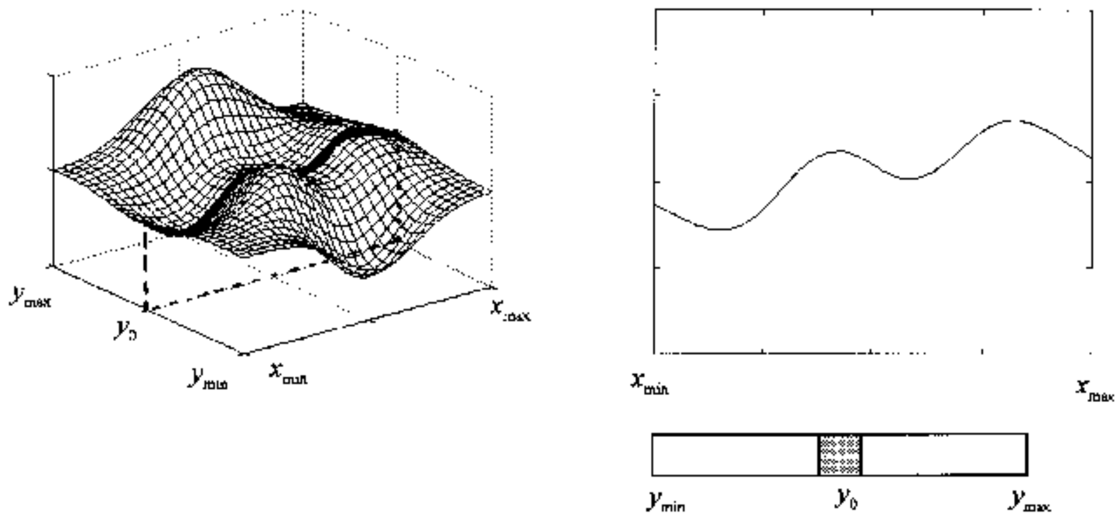


Figure 6. *Schematic illustration of the opportunity offered by fast RBF models to visualize functions of several variables. We see here the concept in the case of a 2-D function visualized by means of 1-D functions. The generalization of d-D functions visualized as 2-D functions is explained in Section 6.2.*

7. Parallel computing.

Massively large supercomputers have received significant attention in recent years, with frequent listings published of the largest computer systems in the world. At the other end of the scale in computing - low-cost PC-type systems - there are present developments that potentially are even more interesting in terms of increased hardware performance, and which are far easier to fully utilize by individual researchers (as well as by teams).

From typically having one single processor, even notebook PCs nowadays usually have two independent processing cores on their main processor chip (then described as a dual core chip), often for desktop PCs going up to two quad core processors (8 cores in all), and soon well beyond this. For example, Intel's recent Nehalem-EX processor features up to eight complete cores on each processor chip, and a PC can have several of these chips. Furthermore, each core can be multithreaded, effectively doubling the core number. In a separate project, a 48-core chip is at present undergoing testing at Intel.

Only one single (not multi-threaded) processing core was used for the comparisons described in this report. However, extensions to use several cores/threads simultaneously are immediate in typical scientific languages (such as Matlab). By changing only a few lines of code, one trivially speeds up any of the here described Matlab code by the same factor as there are processing cores/threads in the computer. We have confirmed this prediction by running the Matlab code described above (for both of the present test problems, Keystroke Data Entry, and RADAR) on a dual quad core PC, in both cases then running 8 times faster than the numbers we reported above.

Another recent multi-core development that also has received much attention is provided by GPUs (Graphics Processor Units), exemplified for instance by the Tesla and Fermi systems manufactured by NVIDIA. For example, a single Fermi plug-in board for a standard PC costs around $6,000, and features 512 independent cores, accessible through convenient Fortran, C++, and Matlab interfaces (in the case of Matlab known as *Jacket*). Although GPUs offer tremendous opportunities in many areas of scientific computing, their applicability to cognitive modeling is at present very uncertain. Just like for most large parallel supercomputing systems, very high performances tend to be linked to the computational tasks being possible to structure in the form of large matrix operations and only very few conditional or sequential statements. GPUs operate in a data-parallel mode, and do not have nearly as much flexibility as CPUs (Central Processing Units) to run with complete independence from each other. Hence, for the foreseeable future (i.e. the next few years), multicore CPUs (rather than GPUs) will be an interesting option for cheaply and easily further speed up equation-based cognitive modeling codes.

References

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of mind. *Psychological Review, 111*, 1036-1060.

Anderson, J. R., & Lebiere, C. (1998). The atomic components of thought. Hillsdale, NJ: Lawrence Erlbaum Associates

Best, B. J., Gonzalez, C., Young, M. D., Healy, A. F., & Bourne, L. E., Jr. (2007). Modeling automaticity and strategy selection in dynamic visual detection. In *Proceedings of the Sixteenth Conference on Behavior Representation in Modeling and Simulation* (pp. 3-11). Orlando, FL: Simulation Interoperability Standards Organization.

Buck-Gengler, C. J., Raymond, W. D., Healy, A. F., & Bourne, L. E., Jr. (2007). Modeling data entry in IMPRINT. In *Proceedings of the Sixteenth Conference on Behavior Representation in Modeling and Simulation* (pp. 205-206). Orlando, FL: Simulation Interoperability Standards Organization.

Buck-Gengler, C. J., Raymond, W. D., Healy, A. F., & Bourne, L. E., Jr. (2010, March). *Modeling a visual search task with a secondary task in IMPRINT*. In Proceedings of the 19th Annual Conference on Behavior Representation in Modeling and Simulation (BRIMS), Charleston, SC, March 22-25, 2010. Available on CD-ROM

Cipra, B. (2000). The best of the 20th century: Editors name top 10 algorithms. *SIAM News*, *33*, 4.

Dongarra, J., & Sullivan, F. (2000). Top 10 algorithms of the century. *Computing in Science and Engineering*, *2*, 22-23.

Fasshauer, G.E. (2007), Meshfree Approximation Methods with Matlab, *World Scientific Publishing Co*, Singapore.

Fornberg, B., Raymond, W.D., & Best, B.J. (2007). *Model Evaluation - ACT-R, IMPRINT, and Matlab: Timing and complexity comparisons on a keystroke data entry test problem*. MURI Internal report.

Gonzalez, C., Fu, W.-T., Healy, A. F., Kole, J. A., & Bourne, L. E., Jr. (2006). ACT-R models of training data entry skills. In *Proceedings of the Fifteenth Conference on Behavior Representation in Modeling and Simulation* (pp. 101-109). Orlando, FL: Simulation Interoperability Standards Organization.

Gluck, K. A., & Pew, R. W. (2001). Lessons learned and future directions for the AMBR model comparison project. *Proceedings of the 10th Annual Conference on Computer-Generated Forces and Behavior Representation* (pp. 113-121). Orlando, FL: Division of Continuing Education, University of Central Florida.

Gluck, K., Scheutz, M., Gunzelmann, G., Harris, J., & Kershner, J. (2007, March). *Combinatorics meets processing power: Large-scale computational resources for BRIMS*. Paper presented at the 16th Conference on Behavior Representation in Modeling and Simulation (BRIMS), Norfolk, VA.

Hardy, R.L. (1971), Multiquadric equations of topography and other irregular surfaces, *Journal of Geophysical Research*, *76*, 1905-1915.

Healy, A. F., Kole, J. A., Buck-Gengler, C. J., & Bourne, L. E., Jr. (2004). Effects of prolonged work on data entry speed and accuracy. *Journal of Experimental Psychology: Applied*, *10*, 188-199.

Pitt, M. A., & Myung, J. (2002). When a good fit can be bad. *Trends in Cognitive Sciences, 6,* 421-425.

Young, M. D., Healy, A. F., Gonzalez, C., Dutt, V., & Bourne, L. E., Jr. (in press). Effects of training with added difficulties on RADAR detection. *Applied Cognitive Psychology*.

Young, M. J. (2003). Human performance model validation: One size does not fit all. *SCSC '03,* 732-736.

The Instance-Based Learning Modeling Tool

Tech Report

Cleotilde Gonzalez
Dynamic Decision Making Laboratory
Carnegie Mellon University

Version 1.1.93, April 12, 2010

Please send requests for errata to the author.

Table of Contents

Chapter 1: Overview **3**
Chapter 2: Introduction **4**
2.1 What is the Instance-based Learning Theory? 4
2.2 What is the Instance-based Learning tool? 5
Chapter 3: Getting Acquainted with the Tool **7**
3.1 Installing 7
3.2 User Interface 7
3.3 Database 8
3.4 Formulas 8
3.4.1 Formula Components 9
3.4.2 Function Calls 11
Chapter 4: Steps to Modeling with Simon Task Example **14**
4.1 Simon Task 14
4.2 Defining Instance Types 15
4.3 Pre-populating Instances into the Memory 16
4.4 Defining Similarity Formulas 18
4.5 Specifying a Match Request 19
4.6 Choosing a Retrieval Method 20
4.7 Setting Judgment Heuristics 20
4.8 Defining Decision-Calculation Formulas 21
4.9 Defining Feedback Formulas 22
4.10 Selecting a Utility Update Method 23
4.11 Setting Model Parameters 24
4.12 Executing the Model 26
Chapter 5: Protocol Definition **28**
5.1 Protocol Format 28
5.2 CUE Message 29
5.3 CUESIZE Message 29
5.4 DECISION Message 29
5.5 ERROR Message 29
5.6 FEEDBACK Message 30
5.7 FEEDBACKOK Message 30
5.8 START Message 30
5.9 STOP Message 30
5.10 STATE Message 31
5.11 Message Flow 31
5.12 Example Message Flow 31
5.13 Planned Changes to the Protocol 32
5.13.1 BATCH Message 32
5.13.2 RESET Message 32
Bibliography **33**

Chapter 1: Overview

This report contains information on how to use the Instance-based Learning tool (IBLtool). The document is written to explain the IBLtool to beginners in modeling techniques as well as to advanced users of modeling and instance-based learning.

Chapter 2 serves as a short introduction to the tool, the theory behind it, and the goals of this tool.

Chapter 3 contains an overview of the tool and its interface.

Chapter 4 takes the Modeler through the steps necessary to create a working model from the beginning to end.

Chapter 5 describes the protocol necessary to connect a task to the tool.

Chapter 2: Introduction

2.1 What is the Instance-based Learning Theory?

The Instance-based Learning Theory (IBLT) was initially proposed to demonstrate how learning occurs in dynamic decision-making tasks (Gonzalez et al., 2003). An IBLT model was implemented within the ACT-R architecture (Anderson and Lebiere, 1998), and we demonstrated how IBLT parameters were needed to account for human decision making in a dynamic and complex task. IBLT has more recently been used in other tasks in addition to dynamic decision making. These include simple binary choice tasks and two-person game-theory learning (Gonzalez & Lebiere, 2005).

Under the IBLT (See Figure 2.1), modelers determine the representation of declarative knowledge (chunks or instances) in a task. In IBLT, an instance is a triple containing the cues that define a situation (S), the actions that define a decision (D), and the expected or experienced value resulting from an action in such a situation (U). Simply put, an instance is a concrete representation of the experience that a human acquires in terms of the decision-making situation encountered by the human, the decision the human makes, and the outcome (feedback) the human obtains.

A modeler following the IBLT approach must define the structure of an SDU instance. Then, an ACT-R modeler following the IBLT approach should define productions that represent the generic decision-making and problem-solving process proposed by IBLT. This process involves the following steps:

- Recognition, the comparison of cues from the environment or task to cues from memory;
- Judgment, the calculation of the possible utility of a decision in a situation, either from past memory of from heuristics;
- Choice, the selection of the instance containing the highest utility; and
- Feedback, the modification of the expected utility defined in the judgment process with the experienced utility after receiving the outcome from a decision made.

The IBLT mechanisms involve a set of functions and thresholds, including a similarity function used in the recognition step to determine what instances from memory are similar to the current situation; the decision threshold used in the choice step to determine whether more “evidence” or alternative search is needed before a selection is made; and the feedback threshold used to determine “how much” of the outcome provided from the environment is accounted for in the utility of the instances.

Instance An *instance* is the smallest unit of an experience. It is a set of values that represent a specific state, which is expressed in a triplet consisting of the Situation, Decision, and Utility slots, or SDU.

Instance Type An *instance type* is a collection of instances with the same structure of the triplet. An instance type may contain more than one of each: situation, decision, and utility slots.

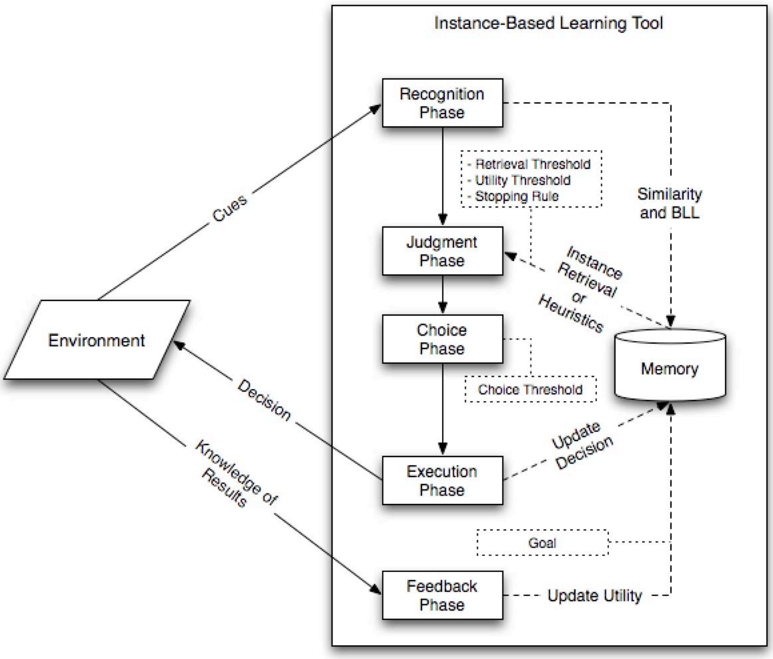


Figure 2.1: Instance-based Learning Theory

2.2 *What is the Instance-based Learning tool?*

The Instance-based Learning tool (IBLtool) is an effort by the Dynamic Decision Making Laboratory to formalize the theoretical approach to modeling. The goals are to have the Instance-based Learning Theory be:

Shareable: by bringing the theory closer to the users, and making it more accessible;

Generalizable: by making it possible to use the theory on different and a diverse set of tasks;

Understandable: by making the theory easier to implement and use;

Robust: by abstracting the specifics of the implementation of the theory away from any specific programming language;

Communicable: by making the tool interact more easily and in a more standard way with tasks; and

Usable: by making the theory more transparent to users.

The tool is a graphical interface written in Visual Basic that uses sockets to communicate with various tasks.

Chapter 3: Getting Acquainted with the Tool

In this chapter, we will get acquainted with the user interface of the IBLtool, and get started with the basic concepts that will help you as you move through the modeling process.

3.1 Installing

To use the IBLtool on your computer, you will need a few things:

1. a Windows XP or Windows Vista machine, with the latest software updates; and
2. the installer package for the tool. There are separate installers for Windows XP and Windows Vista, so ensure you have the correct installer; the files should be named iblt-#.#-xp.exe for Windows XP and iblt-#.#-vista.exe for Windows Vista, where #.# is the version number of the IBLtool.

To install the tool, simply double-click the installer and follow the instructions.

When upgrading the tool, it is recommended that you uninstall previous versions of the software before installing the new version.

3.2 User Interface

The tool is presented as a graphical user interface. It is arranged into successive screens. One such screen can be seen in Figure 3.1.

Each screen is divided into three areas:

Instructions Each screen shows a short set of instructions for actions pertinent to the screen. Instructions appear at the top of the screen.

Content The bulk of a screen's functionality, or content, appears in the middle of the screen. Most screens have a tabbed interface, in which each tab in the tabbed interface represents an instance type. The tabbed interface aims to separate each instance type and reduce confusion as to which instance type is currently being worked on.

Buttons At the bottom of every screen is a collection of buttons. The left-most and right-most buttons are navigation buttons and can be used to move to the previous and next screens, respectively.

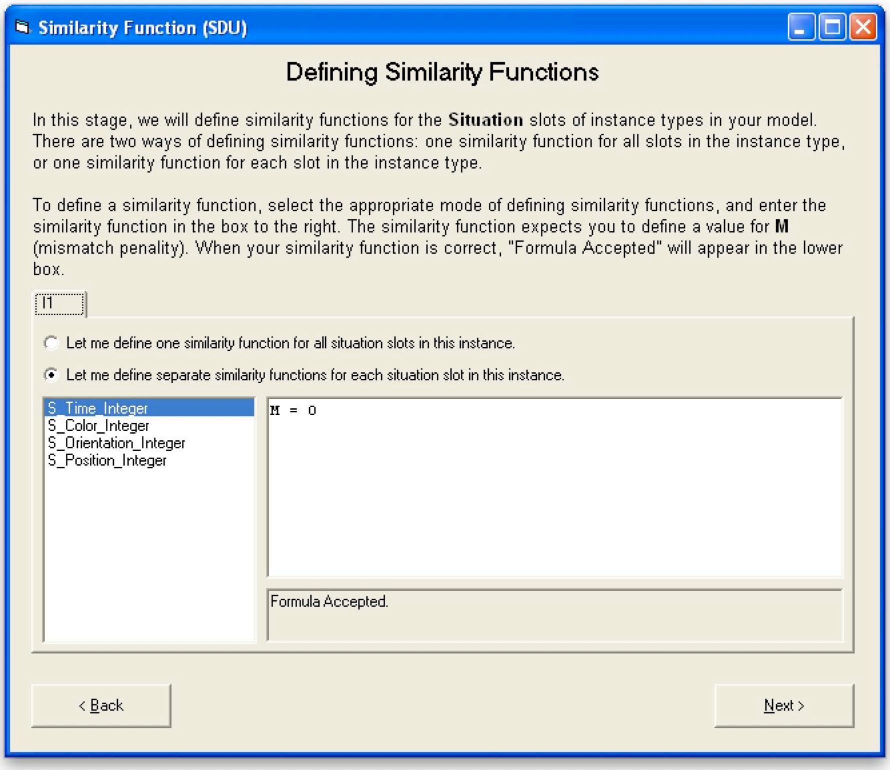


Figure 3.1: An example of a screen in the tool.

3.3 Database

All your instance types, instances, model parameters, and formulas are automatically saved into a database file named instances.mdb.

The tool will create a new, empty database for you when you first start it. To move your model between computers, copy the database file to another computer. Be sure to install the tool on both computers.

The database file can be opened using a copy of Microsoft Access, which can be useful when post-processing data collected during simulation. While it is also possible to modify tool parameters directly from Microsoft Access, we strongly recommend doing so through the tool instead, to prevent the possibility of corrupting any configuration parameters.

3.4 Formulas

Formulas and formula editors are a large part of the tool, because they allow users to write their own formulas using simple arithmetical operations. Formula editors are divided into three sections:

Formula Entry The Formula Entry box is where the users will enter their formula.

Variable List The Variable List box shows a list of variables available for use in that formula. Clicking a variable in the Variable List will insert that variable in the Formula Entry box.

Formula Status The Formula Status box shows whether there are any errors in the formula, if the tool expects the formula to define a certain variable, or if the formula was accepted without any errors.

One important point to note is that formulas written in the formula editor will be automatically checked for errors, and automatically saved.

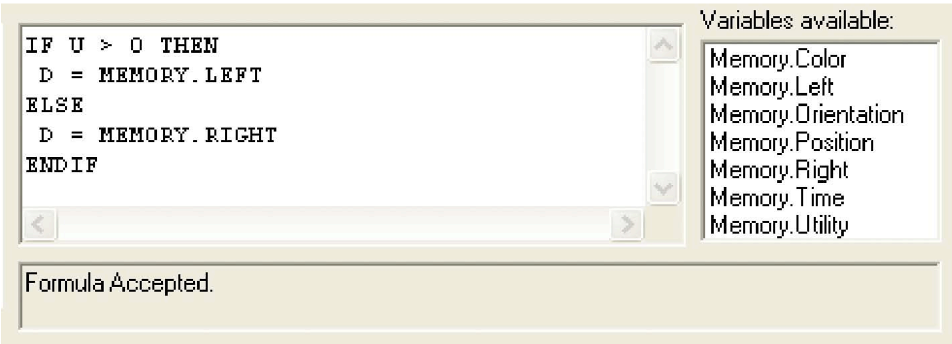


Figure 3.2: Formula Editor, consisting of *Formula Entry* (top left), *Variable List* (top right), and *Formula Status* (bottom). In this example, the formula has been successfully accepted by the tool, i.e. the formula has no errors and all the variables are correctly defined.

3.4.1 Formula Components

Formulas and all their contents—including variables—are case insensitive, i.e. abc is equivalent to ABC. This case-insensitivity will prevent many errors.

Formula A formula consists of one or more Statements. Each Statement must appear on its own line.

Statement A statement can be: *(a)* an Assignment, or *(b)* an IF Conditional.

Assignment An assignment is used to assign a value—or another variable— to a variable. Variable names must start with a letter, but may be followed by any alpha-numeric character (A-Z and 0-9) or a period. For example, these are valid variable names: A, A6, MEMORY, MEMORY.GOAL.

Example formula:

```
A = 5
B = A
```

The formula above consists of two statements, both of which are assignments. When the formula is run, as expected, both A and B will carry the value 5.

IF **Conditional** An IF conditional is used to perform different tasks depending on a set of conditions.

The syntax for IF conditional is:

```
IF condition THEN
     statement1
ELSE
     statement2
ENDIF
```

The condition above is an Expression. Both statement1 and statement2 are regular statements, which would allow the user to have complex rules and nested conditionals.

Expression An expression may be:

1. a variable or value, e.g. TIME or 5;
2. a function call, e.g. ABS(CUE.GOAL);
3. a mathematical computation, e.g. TIME + 5, which uses a mathematical operator (see Mathematical Operators);
4. a comparison, e.g. TIME + 5 > 10, which uses a comparison operator (see Comparison Operators); or
5. a logical expression, e.g. (TIME + 5 > 10) AND (GOAL < 6), which uses a logical operator (see Logical Operators), and connects other expressions together.

Operator	Description	Example
+	Addition	5 + 2
−	Subtraction	CUE.GOAL - MEMORY.GOAL
∗	Multiplication	A * B
/	Division	A / B
\	Division with rounding down	A \ B
∗∗	Exponentiation	2 ** B

Table 3.1: Mathematical operators and their examples.

Operator	Description	Example
==	Equality	MEMORY.TIME == CUE.TIME
<> or ! =	Inequality	MEMORY.TIME <> CUE.TIME
>	Greater than	A > -5
>=	Greater than or equal to	A >= -5
<	Less than	A + B < 2 * A
<=	Less than or equal to	A + B <= 2 * A

Table 3.2: Comparison operators and their examples.

Function call A function call is used to invoke one of the predefined functions in

the tool; it uses the function call operator (), and takes arguments. Each argument is separated by a comma, and an argument is simply any valid expression. For example:

```
Q = ABS(NOISE)
```

calculates the absolute value of the variable NOISE and saves the result into variable Q. The function name in this case is ABS, and it has one argument, denoted by (NOISE).

3.4.2 Function Calls

The IBLtool has various function calls available for use:

ABS(expr1) This function expects one argument, and computes the **absolute value** of that argument.

AVG(expr1, expr2, ..., exprN) This function expects at least one argument, and computes the **mean** value of all arguments.

IIF(expr, exprT, exprF)
The “Immediate IF” function, which is the function-call equivalent of the IF conditional expression, expects three arguments:

expr: the expression to test;
exprT: the expression to use when expr evaluates to TRUE; and
exprF: the expression to use when expr evaluates to FALSE.

Although functionally equivalent to the IF conditional expression, the IIF function has a limitation that comes from the fact that it can only process expressions, and not statements.
Compare the IF conditional:

```
IF MEMORY.GOAL < CUE.GOAL THEN
      DECISION = 0
ELSE
      DECISION = MEMORY.GOAL - CUE.GOAL
ENDIF
```

to the IIF function-call (formula broken into two lines due to length):

```
DECISION = IIF(MEMORY.GOAL < CUE.GOAL, 0,
      MEMORY.GOAL - CUE.GOAL)
```

In this case, the above two examples are equivalent: they will set DECISION to 0 if MEMORY.GOAL is less than CUE.GOAL, and set DECISION to the difference otherwise.

To illustrate the limitation of IIF, consider the conditional:

```
IF MEMORY.GOAL < CUE.GOAL THEN
      LEFT = 1
      RIGHT = 0
ELSE
      LEFT = 0
      RIGHT = 1
ENDIF
```

In this case, the IF conditional cannot be expressed as an IIF function call.

LOG(expr, exprBase) This function expects two arguments, and computes the base exprBase logarithm of expr.

LN(expr) This function expects one argument, and computes the natural logarithm of expr.

MAX(expr1, expr2, ..., exprN) This function expects at least one argument, and computes the **maximum** of all arguments.

MIN(expr1, expr2, ..., exprN) This function expects at least one argument, and computes the **minimum** of all arguments.

POWER(exprBase, exprExponent) This function expects two arguments: the base number (exprBase) and the exponent number (exprExponent).

RAND() **or** RAND(expMax) **or** RAND(expMin, expMax)

This function expects no, one, or two arguments, and returns a **randomly-generated number**.

- When called with no argument, it returns a number between 0 and 1.
- When called with one argument, it returns a number between 0 and expMax.
- When called with two arguments, it returns a number between expMin and expMax.

RANDITEM(expr1, expr2, ..., exprN)

This function expects at least one argument, and **randomly chooses** one of the supplied arguments. Each argument has equal probability of being selected. For example, the following formula randomly chooses between the value of MEMORY.GOAL and the value of CUE.GOAL:

```
DECISION = RANDITEM(MEMORY.GOAL, CUE.GOAL)
```

ROUND(expr)

This function expects one argument, and returns **the Gaussian rounding** of the value passed to it; i.e. fractional values are rounded to the nearest even integer. For example: both 15.5 and 16.5 are both rounded to 16. Gaussian rounding is the rounding implementation used by Visual Basic.

SQRT(expr) This function expects one argument, and computes the **square-root** of the argument. It is essentially equivalent to expr ** 0.5.

SUM(expr1, expr2, ..., exprN) This function expects at least one argument, and computes the **sum** of all arguments.

Chapter 4: Steps to Modeling with Simon Task Example

This chapter will cover the steps needed to model a task using the IBLtool.

Before starting, there are a few points to remember:

1. You do **not** need to have the task running to begin modeling.
2. You need both the task program and the tool installed to perform simulations. They may be installed on the same or different computers. If they are on different computers, it is highly suggested that both computers be on the same local computer network to reduce the possibility of network latency issues. Network latency issues may cause the task or the tool to fall behind from one or the other, and cause problems with your simulations.
3. The task to which you are using must be modified—if not already— to be able to connect to the tool. Your developer—or the person who originally wrote the task program you are using—can refer to Protocol Definition for information on what changes are needed.

This is both a guide and tutorial, so each step will relate back to an example task, the Simon Task, which will be reviewed in the next section.

4.1 Simon Task

First, let us run through a brief overview of the task we will be using: the Simon Task.

The Simon task is a location-irrelevant choice-reaction task. In the task, subjects are shown stimuli in the form of five-millimeter red or green circles on the screen. Responses are made by pressing one of two keys: a left key, or a right key. When a red circle is shown, one response key must be pressed, while when a green circle is shown, the other response key must be pressed.

Because the Simon task is location irrelevant, the same key must be pressed every time the same-colored circle appears, regardless of where the circles appear on the screen.

Each trial starts with a white fixation cross at the center of the screen for 500 milliseconds, followed by a blank screen for 500 milliseconds, and finally a red or green circle is shown on the left- or right-side of the screen. Subjects have up to 1,500 milliseconds to provide a response—correctly or incorrectly. Incorrect responses produce an error tone, while no feedback is given for a correct response.

4.2 *Defining Instance Types*

The first step is to define the structure of one or more instance types. Most tasks will have one instance type, but the tool supports having multiple instance types.

From the description of the task above, we can construct the following instance type:

Situation (S)				Decision (D)		Utility (U)
Time	Color	Orientation	Position	Left	Right	Utility
.	.	.	.	.	.	.

All the situation and decision slots are integer value, while the utility slot is a floating or real value. In the above example, all slots are empty (·).

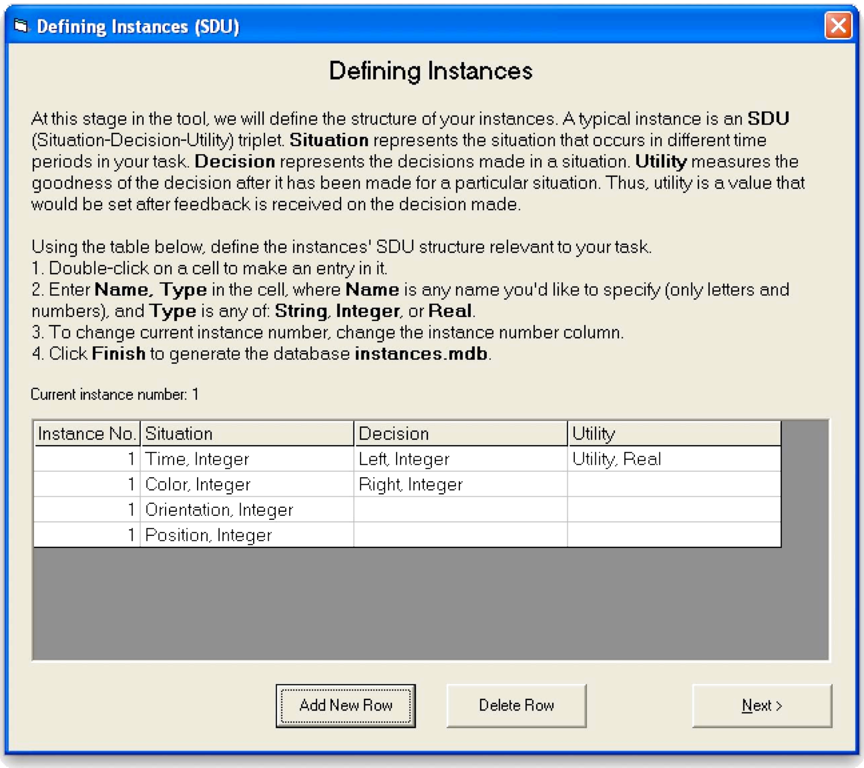
Because the color, orientation, and position situation slots are categorical but stored as integers, it is recommended that a coding table that maps the integer values to actual value is kept for your reference. For example:

Slot	Code	Actual Value
Color	0	Green
	1	Red
Orientation	0	Horizontal
	1	Vertical
Position	0	Left
	1	Right

You can construct and modify instance types on the first screen of the tool.

To add a new slot on the instance:

1. Click the **Add New Row** button.
2. Double-click the slot which you would like to add.
3. Type the slot name, followed by a comma, followed by the type of slot.
4. Press enter to add the slot, or escape to cancel the addition.



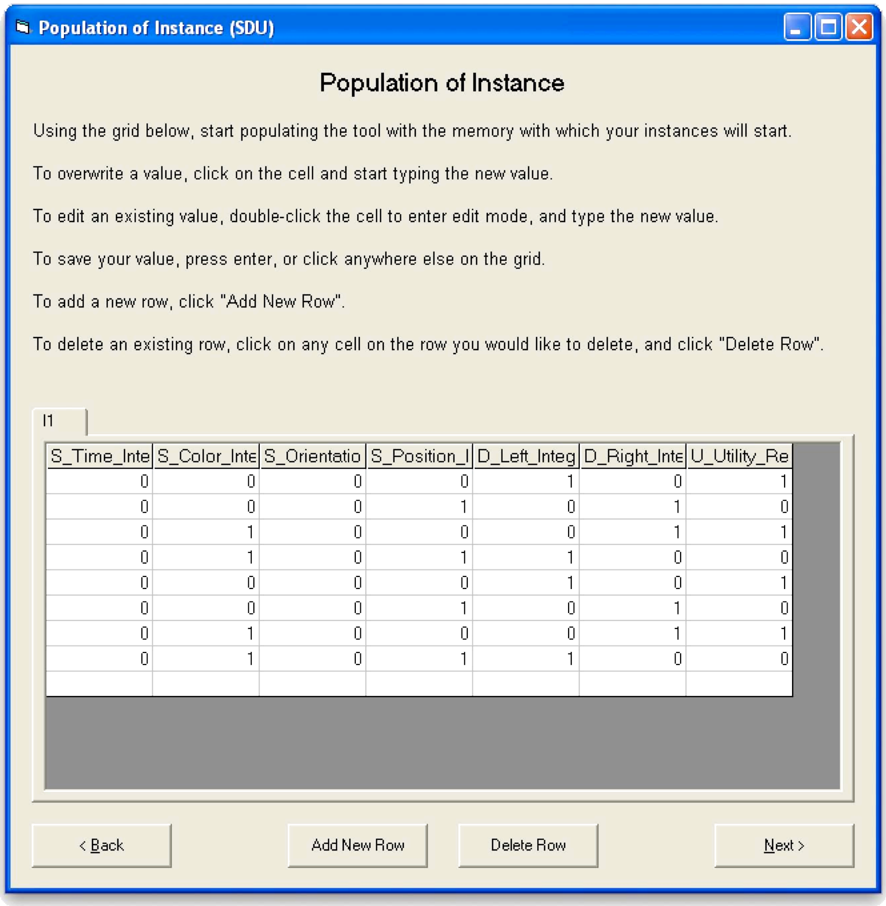
For example, to add the **Time** situation slot as an integer, we would type **Time, Integer**.

The tool currently supports three types of values: **Integer**, **Real**, and **String**. To store categorical values, it is recommended to assign each possible value to a numerical value and use Integer fields instead of String fields.

4.3 Pre-populating Instances into the Memory

Next, we can start pre-populating the tool's memory with instances. This step is *completely optional*, and can be safely skipped.

When a simulation starts, pre-populated instances will be treated as if they were added at the very start of the simulation.



To add a new instance to the memory:

1. Click the **Add New Row** button.
2. Double click the first cell on the new row, and start entering the value.
3. Press enter to save a value, or esc to cancel adding the value. When you press enter, the next cell—if any—will be automatically editable. This allows you to quickly add instances without having to use the mouse.

To delete an instance from the memory:

1. Click on any cell on the row which you would like to delete.
2. Click the **Delete Row** button.

To edit an existing instance:

1. Click on the cell of the instance you would like to edit.
2. Enter the new value.
3. Press enter to save, or esc to cancel the edit.

For the purposes of our example, we have added the following pre-populated instances into memory, which is every possible combination of color, position, and

answer:

S				D		U
Time	Color	Orientation	Position	Left	Right	Utility
0	0	0	0	1	0	1
0	0	0	1	0	1	0
0	1	0	0	0	1	1
0	1	0	1	1	0	0
0	0	0	0	1	0	1
0	0	0	1	0	1	0
0	1	0	0	0	1	1
0	1	0	1	1	0	0

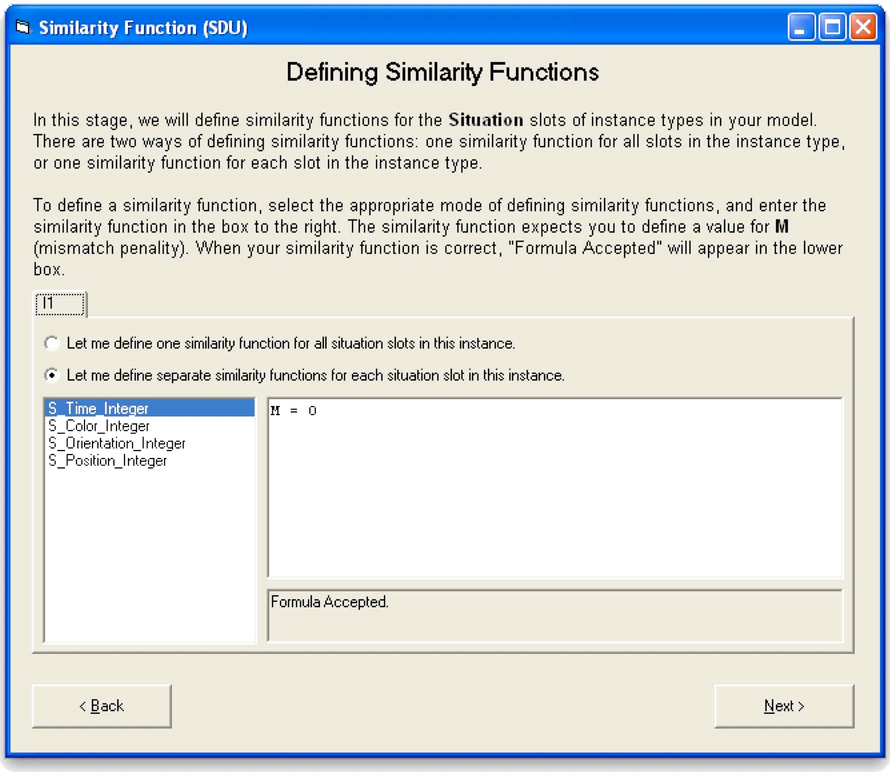
4.4 Defining Similarity Formulas

In this screen, you will see your first formula editor (see Formulas for an introduction to formulas), in which you will be able to specify one or more similarity formulas. Similarity formulas can only be defined on situation slots.

There are currently two ways of specifying similarity functions:

- **Define one similarity formula for all slots**
  When this option is selected, you will be able to enter a formula for calculating similarity into the formula editor, which will then be used to calculate similarity for every situation slot within that instance type.
- **Define a separate similarity formula for each slot**
  When this option is selected, the sidebar will activate and allow you to select a situation slot for which to define a similarity formula. To start adding a similarity formula, click on a slot name and start writing the formula.

The formula editors on this screen expect you to define the variable M (mismatch penalty).



For the purposes of our example, we have defined separate similarity formulas for each slot:

Slot	Formula
Time	M = 0
Color	M = -1 * ABS(CUE - MEMORY)
Orientation	M = -1 * ABS(CUE - MEMORY)
Position	M = -1 * ABS(CUE - MEMORY)

4.5 Specifying a Match Request

Currently, during the retrieval process, all instances in memory are candidates for retrieval.

In some tasks however, this may not be the desirable course of action. As such, in this screen, you have the opportunity to limit retrieval only to instances in memory that satisfy certain criteria.

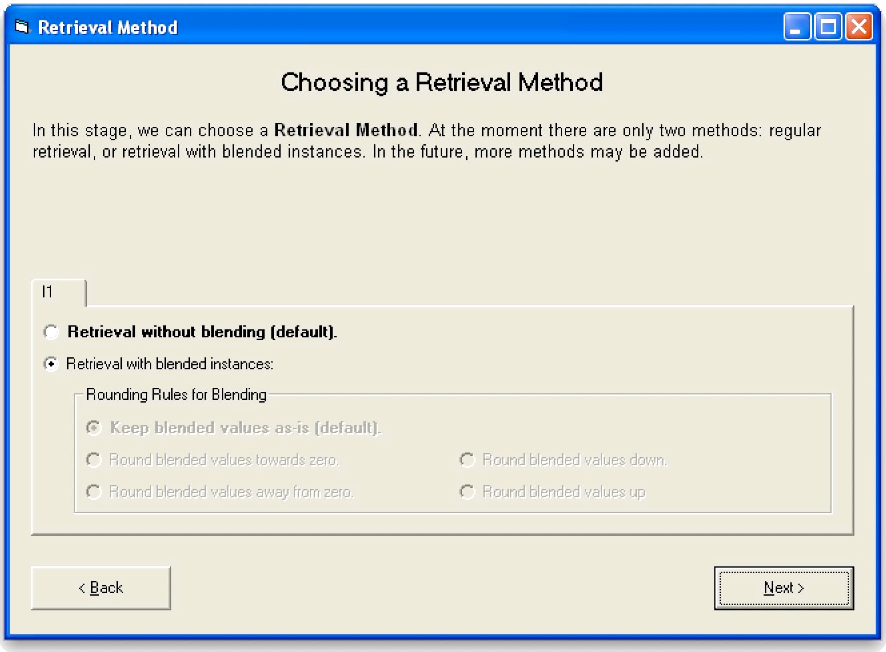
For the purposes of our example, we have selected to only take into account instances of the same color as the cue, regardless of the utility of said instance or any other slot value:

```
IF  CUE.COLOR  ==  MEMORY.COLOR  THEN
     USES  =  TRUE
ELSE
     USES  =  FALSE
ENDIF
```

4.6 Choosing a Retrieval Method

In this screen, you can choose the retrieval method you would like to use. There are currently two options:

- **Regular retrieval**
  In regular retrieval, instances are first marked as candidate for retrieval if they fulfill the Match Request. Of those instances that are candidates, the instances with the best activation score that satisfy the Request Threshold and Utility Threshold—if any such instances exist—will be retrieved; otherwise, retrieval will fail.
- **Retrieval with blended instances**
  In retrieval with blended instances, instances are also first marked as candidate for retrieval if they fulfill the Match Request. If there is at least one candidate instance, the retrieval process will create a new chunk of the same instance type, whose slots are the blended values of all the candidate instances. If there are no candidate instances, retrieval will fail.



For the purposes of our example, we have selected to use blended instances.

4.7 Setting Judgment Heuristics

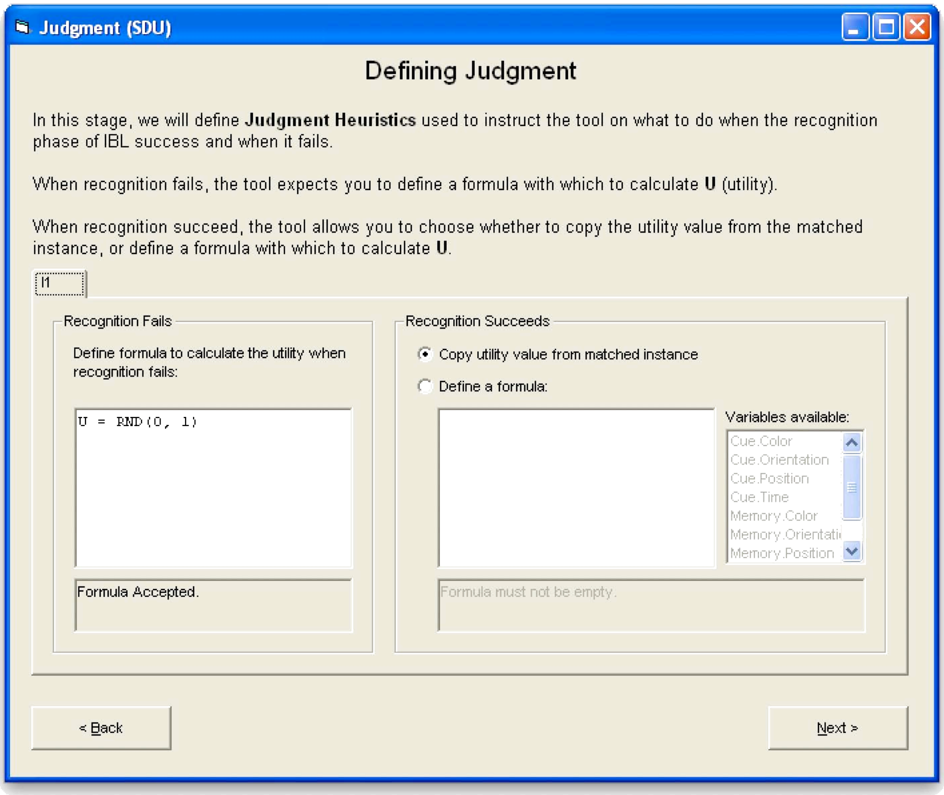
In this screen, you will have the chance to define judgment heuristics. After retrieval is performed, the tool will either succeed in retrieval, in which case an instance was retrieved, or fail, in which case no instance was retrieved.

When retrieval fails, the tool expects you to define a formula to calculate the

utility value. The formula expects you to define the variable U (expected utility value).

When retrieval succeeds, there are two choices:

- **Copy utility**
  The utility value can be copied from the instance that was retrieved.
- **Utility formula**
  The utility value can be calculated based on a formula. The formula expects you to define the variable U (expected utility value). The formula will have access to all the slot values of the cue that triggered the retrieval, and the instance that was retrieved.



For our example, we will simply copy the utility value upon successful retrieval. We will also define the following formula to calculate the utility value upon failed retrieval, essentially assigning the utility a random value between 0 and 1:

```
U = RAND(0, 1)
```

4.8 Defining Decision-Calculation Formulas

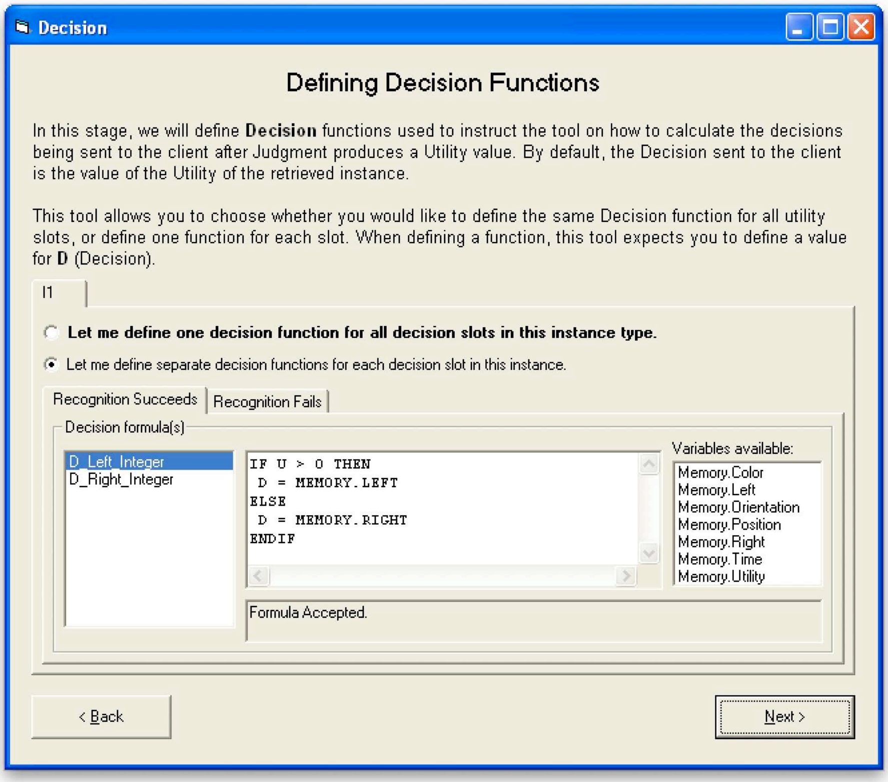
In this screen, you can define how a decision value is calculated, and sent back.

There are two options when defining decision calculation formulas:

- **Define one decision formula for all decision slots**
  When this option is selected, you will be able to enter a formula into the formula editor, which will then be used to calculate similarity for every decision slot within that instance type.

- **Define a separate decision formula for each decision slot**
  When this option is selected, the sidebar will activate and allow you to select a decision slot for which to define a formula. To start adding a similarity formula, click on a slot name and start writing the formula.

Each decision formula expects you to define the variable D (decision value). Furthermore, the tool allows you to define a separate decision formula depending on whether retrieval succeeded or failed.



For the purposes of our example, we have defined separate decision formulas for each slot:

Retrieval	Slot	Formula
Succeed	Left	D = IIF(U > 0, MEMORY.LEFT, MEMORY.RIGHT)
	Right	D = IIF(U > 0, MEMORY.RIGHT, MEMORY.LEFT)
Failed	Left	D = IIF(U > 0.5, 0, 1)
	Right	D = IIF(U > 0.5, 1, 0)

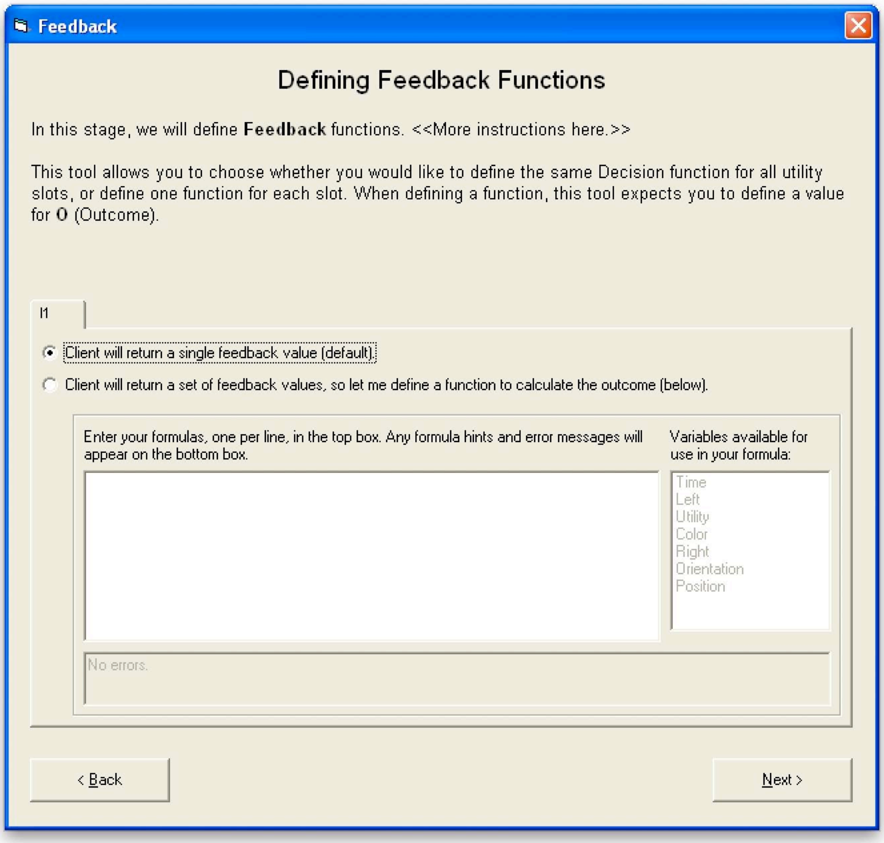
4.9 Defining Feedback Formulas

In this screen, you can define how the tool will process incoming feedback from the task. There are two options available to you:

- **Single feedback value**
  When this option is selected, the tool will expect the task to send a single

value as its feedback. This single value will be used as the value of O (the outcome).

- **Multiple feedback values**
  When this option is selected, the tool will expect the task to send multiple values in one feedback. You will be able to define a formula to calculate the value of O (the outcome) based on the fields in the feedback.



For our example, we will select a single feedback value.

4.10 *Selecting a Utility Update Method*

In this screen, we will use the O (outcome), G (goal, which is a model parameter), and U (expected utility value) to calculate U’ (experimental utility value).
There are three options available:

- **Increase the utility by the outcome**
  When this option is selected, the experimental utility value will be increased based on the outcome value, scaled by the goal value. In other words:

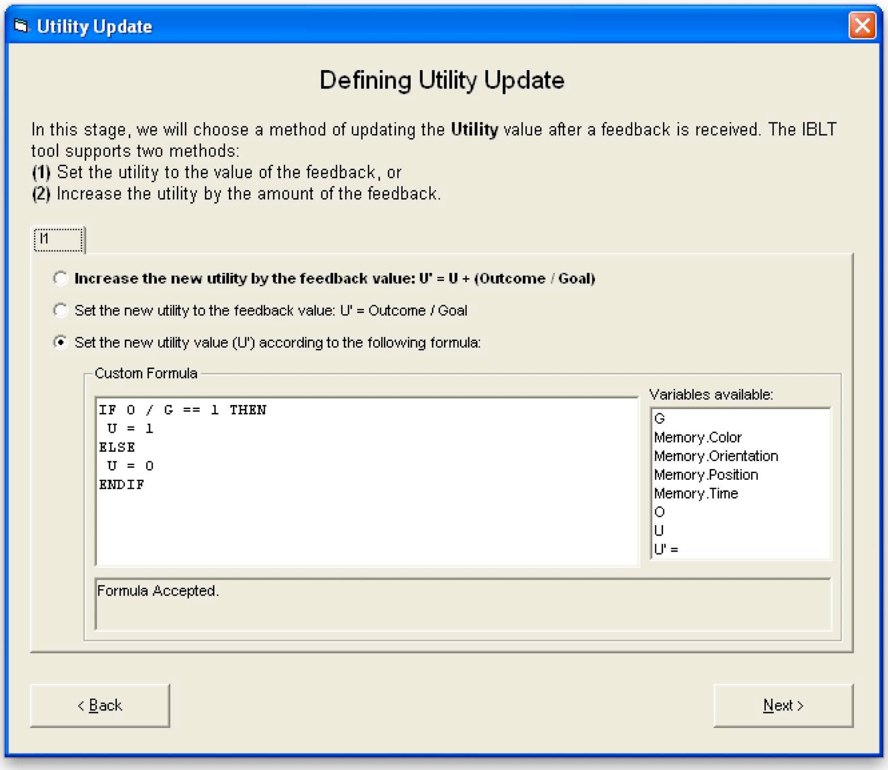
```
U’ = U + (O / G)
```

- **Set the utility to the outcome**
  When this option is selected, the experimental utility value will be set to the

outcome value, scaled by the goal value. In other words:

```
U′ = O / G
```

- **Define a custom formula**
  When this option is selected, you will have the opportunity to enter a custom formula to calculate the experimental utility value.



For the purposes of our example, we will define a custom formula:

```
IF O / G == 1 THEN
  U′ = 1
ELSE
  U′ = 0
ENDIF
```

4.11 Setting Model Parameters

In this screen, you will have the opportunity to specify various model parameters. The model parameters are divided into three areas:

Stopping Rules

All the stopping rule parameters are grouped to the left-hand side of the screen. These parameters include:

- RT (Retrieval Threshold);
- UT (Utility Threshold);

- IBLT Cycle Threshold, for which there is the ability to specify a time-based threshold or a number-of-retrieval threshold;
- CT (Choice Threshold); and
- G (Goal).

Activation-Calculation Parameters

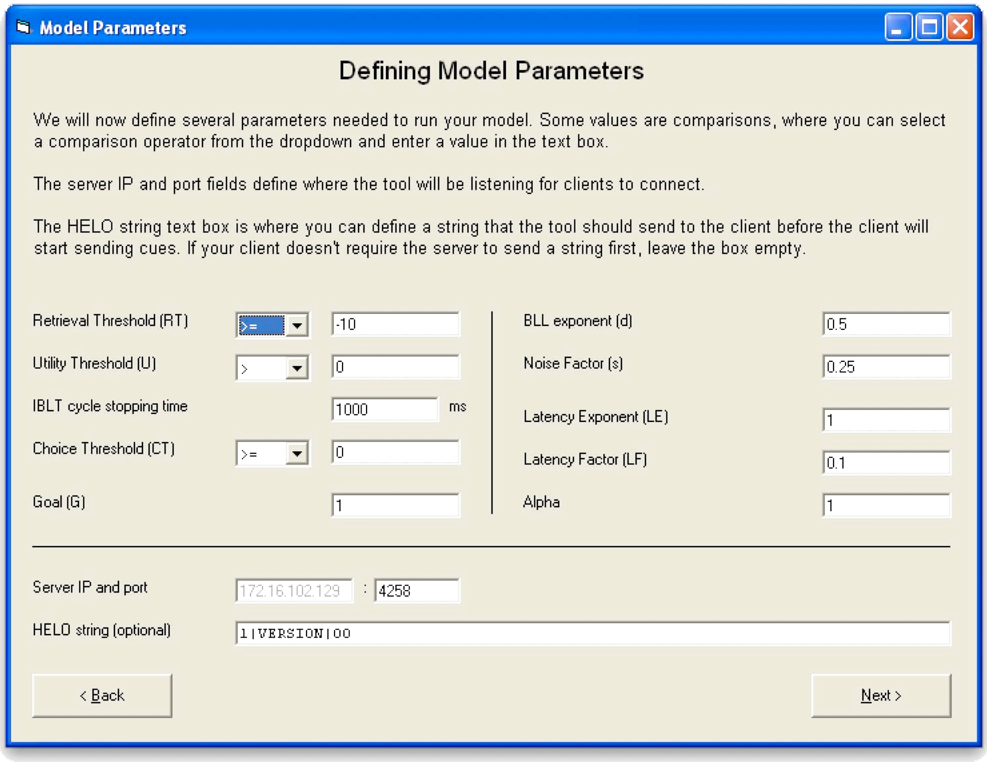
All the parameters that are used when calculating instance activation are grouped to the right-hand side of the screen. These parameters include:

- d, which is the Base-Level Learning Exponent;
- s, which is the Noise Factor;
- LE (Latency Exponent);
- LF (Latency Factor); and
- Alpha, or .

Socket Parameters

The tool interacts with tasks through a network programming—or socket—interface. To control this interface, the tool also comes with additional parameters:

- Server IP, which is the IP address to which the task should connect, and is not a configurable parameter;
- Server Port, which is the port number to which the task should connect; and
- HELO String, which is an optional and configurable string that the tool sends to the task when the first connection is made.

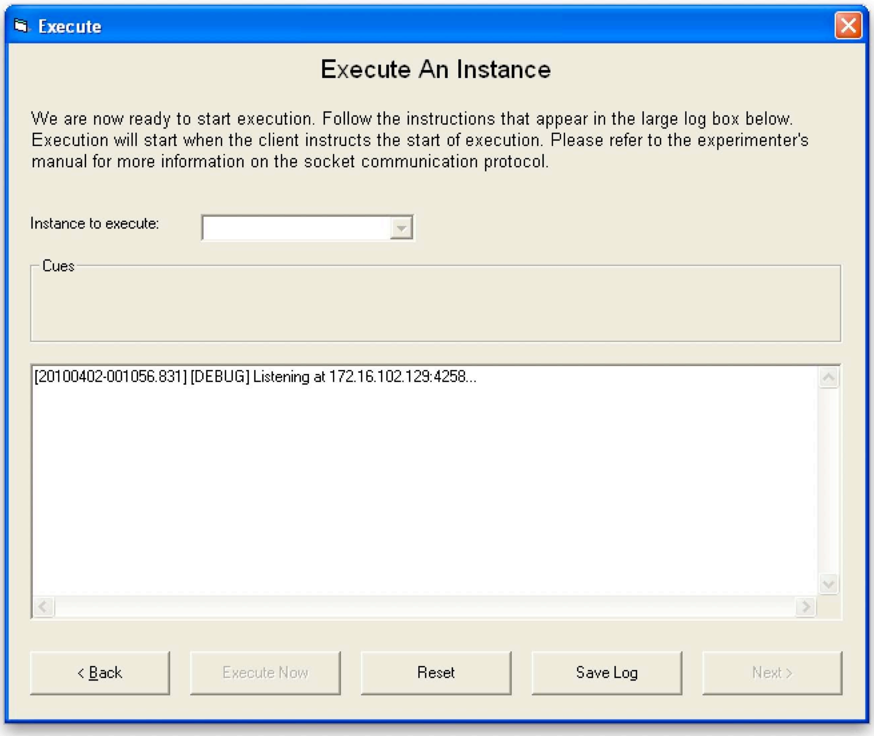


For the purposes of our example, we will use the following parameters:

Parameter	**Setting**
RT	>= −10
UT	>= 0
Cycle Rule	Number of Retrievals: 1
CT	>= 0
G	1
d	0.5
s	0.25
LE	1
LF	0.1
Alpha	1
Port	4258
HELO String	1|VERSION|00

4.12 Executing the Model

In this screen, you will finally have the chance to run the simulation. When you first arrive at this screen, the tool should show a message that it is listening for a connection, and ready to perform a simulation. When this happens, you can start a simulation.



To start a simulation:

1. Start up your task.
2. Connect your task to the tool, and the simulation should commence shortly thereafter.
3. If your task has a **batch mode** and is running in batch mode, then the next simulation will begin as soon as the current one ends.

To reset a simulation when your task is in **batch mode**, click the Reset Simulation button.

To reset a simulation when your task is in **regular mode** or if your task does not have a batch mode:

1. Stop your task in order to stop the simulation in the tool.
2. Click the Reset Simulation button.
3. Start your task back up.

Chapter 5: Protocol Definition

This chapter documents the protocol used by the IBLtool to communicate with a task. You may skip this chapter if:

- the task to which you are connecting has already been modified to connect to the tool; or
- you are only using the tool to create models, and someone else is in charge of modifying your task to connect to the tool.

5.1 Protocol Format

The IBLtool uses a line-based protocol, i.e. each message appears on its own line, and each line is always terminated by \r\n (a carriage return and a new-line character).

There are nine types of messages, each of which will be described in detail in this chapter.

message → *cue* | *cue-size* | *decision* | *error*
| *feedback* | *feedback-ok* | *state* | *start* | *stop*

crlf → "\r\n"

A message consists of one or more fields. Each field is separated by | (the vertical bar, or pipe character).

sep → "|"

Numerical values are either integers or reals, both signed and unsigned.

sign → "+" | "-"
digits → *digit* | *digit digits*
integer → *digits* | *sign digits*
real → *digits* "." *digits* | *sign digits* "." *digits*

String values for our purpose are the list of all printable characters except the terminator and separator.

string-char = printable - sep – crlf
string-chars → *string-char* | *string-char string-chars*
string → *string-chars*

An instance type is occasionally used to denote the instance with which the command is associated. The instance type is simply a string that always starts with "I" followed by numbers.

instance-type → "I" *digits*

Slot values are conveyed using the concept of slot pairs. A slot pair consists of a slot name and a slot value.

slot-name → *string*
slot-value → *real* | *integer* | *string*
slot-pair → *slot-name sep slot-value*
slot-pairs → *slot-pair* | *slot-pair sep slot-pairs*

5.2 CUE Message

The CUE message is used by the task to convey a set of cue values to the tool. A cue is denoted by the "CUE" command followed by the instance type and one or more slot pairs.

cue → "CUE" *sep instance-type sep slot-pairs crlf*

The tool expects the number of slot pairs to coincide with the number returned by CUESIZE Message.

5.3 CUESIZE Message

The CUESIZE message is used to convey the length of cues to expect. It allows the tool to declare a predetermined number of cues to the task.

size → *integer*
cue-size → "CUESIZE" *sep instance-type sep size crlf*

5.4 DECISION Message

The DECISION message is used by the tool to convey one or more decisions back to the task. A decision may either be one single un-annotated value in the event that the task only produces a numerical value, or a list of slot pairs.

single-decision → "DECISION" *sep instance-type sep real crlf*

multi-decision → "DECISION" *sep instance-type sep slot-pairs crlf*

decision → *single-decision* | *multi-decision*

5.5 ERROR Message

The ERROR message is used to convey arbitrary error messages from the tool to the task (but not the other way around).

error-message → *string*
error → *“*ERROR” *sep error-message crlf*

5.6 FEEDBACK Message

The FEEDBACK message is used by the task to send a feedback value into the tool.

feedback-value → *integer* | *real*
feedback → *“*FEEDBACK” *sep instance-type sep feedback-value crlf* |
*“*FEEDBACK” *sep instance-type sep slot-pairs crlf*

Note: Because feedbacks are processed asynchronously, the task can either wait for the FEEDBACKOK message, or ignore FEEDBACKOK altogether if the task doesn’t need to know when feedbacks are processed.

5.7 FEEDBACKOK Message

The FEEDBACKOK message is used by the tool to signal to the task that a feedback has been processed. The acknowledgment also includes the goodness value (goodness-value) applied, and the number of instances to which the feedback was applied (apply-size).

apply-size → *integer*
goodness-value → *integer* | *real*
feedback-ok → *“*FEEDBACKOK” *sep goodness-value sep apply-size crlf*

5.8 START Message

The START message is used by the task to initiate a new simulation on the tool.

start → *“*START” *sep instance-type crlf*

5.9 STOP Message

The STOP message is sent by the task to clean up after a simulation.

stop → *“*STOP” *sep instance-type crlf*

5.10 STATE Message

The STATE message is used by the task to insert a cue and feedback at the same time. The feedback portion will be executed before the cue portion will.

state → “STATE” *sep slot-pairs crlf*

5.11 Message Flow

When starting up, data streams are initiated by the task, not the tool. The general message flow is:

1. Task connects to the tool.
2. Task sends START.
3. Tool sends CUESIZE to the task.
4. Tool starts simulation for the instance type.
5. Task sends CUES or FEEDBACK; tool sends DECISION or FEEDBACKOK.
6. Task sends STOP when it is done.
7. Tool stops simulation for the instance type.
8. Task disconnects.

During simulation, the following events may come in any order:

1. A set of cues (CUES) may come from the task, to which the server will respond with a DECISION.
2. A feedback value (FEEDBACK) may come from the task, to which the server will respond with an acknowledgment (FEEDBACKOK).

5.12 Example Message Flow

Let us assume a simulation is performed on an instance type I2 with 4 situation slots. The C lines denote the task commands sent by the task, while S lines denote the server responses sent by the tool.

The task opens a connection to the tool, and indicates that it wants to perform a simulation on instance type *I*2. The tool informs the task that it will expect four cue (situation) slots.

```
C: START|I2
S: CUESIZE|I2|4
```

The task sends a feedback—even though no cue has been sent—and the tool replies with the feedback value and the number of instances to which the feedback was applied (in this case, none).

```
C: FEEDBACK|I2|60
S: FEEDBACKOK|60|0
```

The task sends a cue to the tool, and the tool sends back a decision value.

```
C: CUES|I2|TIME|1|COLOR|0|POSITION|1|ORIENTATION|1
S: DECISION|85
```

The task sends a feedback to the tool, and this time the tool applies the feedback to one executed instance.

```
C: FEEDBACK|I2|90
S: FEEDBACKOK|90|1
```

The task stops the simulation and disconnects from the tool.

```
C: STOP
```

5.13 Planned Changes to the Protocol

5.13.1 BATCH Message

The BATCH message is used by the task to perform a batch of simulations on the tool, running one simulation after another until the number of requested simulations is performed. The message must be the first message sent to the tool when the task connects.

batch → "BATCH" *sep number-of-simulations crlf*

5.13.2 RESET Message

The RESET message forcefully resets the simulation.

reset → "RESET" *crlf*

Bibliography

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science, 27*(4), 591-635.

Gonzalez, C., & Lebiere, C. (2005). Instance-based cognitive models of decision making. In D. Zizzo & A. Courakis (Eds.), *Transfer of knowledge in economic decision-making* (pp. 148-165). New York: Macmillan (Palgrave Macmillan).

Experimental Research on Training Principles: MURI Research at the University of Colorado

Alice F. Healy and Lyle E. Bourne, Jr.

University of Colorado, Boulder

PRIVILEGED AND CONFIDENTIAL; PLEASE DO NOT QUOTE OR DISTRIBUTE WITHOUT PERMISSION

Publications in Peer-Reviewed Journals

Bonk, W. J., & Healy, A. F. (2010). Learning and memory for sequences of pictures, words, and spatial locations: An exploration of serial position effects. *American Journal of Psychology, 123,* 137-168.

A serial reproduction of order with distractors task was developed to make it possible to observe successive snapshots of the learning process at each serial position. The new task was used to explore the effect of several variables on serial memory performance: stimulus content (words, blanks, and pictures), presentation condition (spatial information vs. none), semantically categorized item clustering (grouped vs. ungrouped), and number of distractors relative to targets (none, equal, double). These encoding and retrieval variables, along with learning attempt number, affected both overall performance levels and the shape of the serial position function, although a large and extensive primacy advantage and a small 1-item recency advantage were found in each case. These results were explained well by a version of the scale-independent memory, perception, and learning model that accounted for improved performance by increasing the value of only a single parameter that reflects reduced interference from distant items.

Bourne, L. E., Jr., Healy, A. F., Bonk, W. J., & Buck-Gengler, C. J. (in press). Intention to respond in a special way offers some protection against forgetting associations. *American Journal of Psychology*.

In a continuous memory-updating paradigm, subjects studied name-color associations and were tested later for the color associate given the name. The default color response was made in one location, but on designated trials the response was required to be made in a special location. Memory for the color associated with a given name was assessed after short and long retention intervals when both default and special responses were required. Separate measures were examined of memory for the intention to respond in a particular way (default or special) and of memory for the color associations paired with the names. Memory for color associates was better overall with short than with long retention intervals and was better when special (rather than default) responses were required, especially at the long retention interval. These results imply that the requirement to respond in a special way protects associations from loss due to forgetting

Bourne, L. E., Jr., Raymond, W. D., & Healy, A. F. (2010). Strategy selection and use during classification skill acquisition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 500-514.

Two experiments examined 3 variables affecting accuracy, response time, and reports of strategy use in a binary classification skill task. In Experiment 1, higher rule cue salience, allowing faster rule application, produced higher aggregate rule use than lower rule cue salience. After participants were pretrained on the relevant classification rule, rule reports were high but generally declined across training trials; after participants were pretrained on an irrelevant rule, reports of

the relevant rule increased across training trials. In Experiment 2, no rule pretraining produced a pattern of results like that obtained with irrelevant rule pretraining in Experiment 1. Presenting novel stimuli during training in Experiment 2 elevated aggregate rule reports relative to conditions where they were absent. Two participant subgroups were identified: those persisting in rule reports and those transitioning from rule to memory reports during training. The proportion of persistent rule users was higher after rule discovery than after relevant rule pretraining. Overall, the results indicate that differences among prior experiments can be reconciled. Further, they raise questions about the inevitability of memory-based automaticity in binary classification, favoring instead strategy choice based on the costs and benefits of a particular strategy and of a shift from one strategy to another.

Healy, A. F., Shea, K. M., Kole, J. A., & Cunningham, T. F. (2008). Position distinctiveness, item familiarity, and presentation frequency affect reconstruction of order in immediate episodic memory. *Journal of Memory and Language, 58,* 746-764.

Three experiments examined the effects of position distinctiveness, item familiarity, and frequency of presentation on serial position functions in a task involving reconstructing the order of a subset of 12 names in a list of 20 names. Three different serial position conditions were compared in which the subset of names occurred in Positions 1–12, 5–16, or 9–20, with all subsets including Positions 9–12. The serial positions were defined temporally in Experiments 1 and 2 and spatially in Experiment 3. The serial position functions in all three experiments were well predicted by Murdock's [Murdock, B. B., Jr. (1960). The distinctiveness of stimuli. Psychological Review, 67, 16–31] account in terms of the distinctiveness of the absolute positions. Experiment 3 also revealed significant effects of item familiarity and frequency of presentation on order reconstruction.

Healy, A. F., Wohldmann, E. L., Sutton, E. M., & Bourne, L. E., Jr. (2006). Specificity effects in training and transfer of speeded responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 534-546.

In 3 experiments, participants, on signal, moved a cursor from a central position to 1 of 8 numerically labeled locations on the circumference of a clock face. Movements were controlled by a mouse in 1 of 4 conditions: vertical reversal, horizontal reversal, combined reversals, or normal (i.e., no reversals). Participants were trained in 1, 2, or 3 of these conditions and were tested 1 week later with either the same or a different condition. There were improvements across training and perfect retention across the delay. There was little or no transfer, however, even when training involved combined reversals or multiple conditions. These results illustrate severe specificity of training and are interpreted in terms of acquired inhibition of normal responses.

Kole, J. A., & Healy, A. F. (2007). The effects of memory set size and information structure on learning and retention. *Psychonomic Bulletin & Review, 14,* 693-698.

Two experiments examined the effects of memory set size and information structure on learning and retention. Participants learned 48 (small set) or 144 (large set) facts about individuals, and were tested over 48 facts. The test facts included either 4 facts about 12 individuals (12-person condition) or 12 facts about 4 individuals (4-person condition). During learning, there was an advantage for the small-set group in the 4-person condition, but a disadvantage in the 12-person condition. During testing, there was an advantage for the 4-person condition relative to the 12-person condition for the small-set group, even when the conditions were equated in terms of name exposure. The results support a mental model account of memory representation and retrieval.

Kole, J. A., & Healy, A. F. (2007). Using prior knowledge to minimize interference when learning large amounts of information. *Memory & Cognition, 35,* 124-137.

In three experiments, we examined mediated learning in situations involving learning a large amount of information. Participants learned 144 "facts" during a learning phase and were tested on facts during a test phase. In Experiments 1 and 2, participants learned facts about familiar individuals, unfamiliar individuals, or unfamiliar individuals associated with familiar individuals. Prior knowledge reduced interference, even when it played only a mediating role. In Experiment 3, participants learned facts about unfamiliar individuals or unfamiliar countries, with half the participants in each group associating the unfamiliar items with familiar individuals. Again, use of prior knowledge to mediate learning reduced interference even when the new information was conceptually dissimilar to the previously known information. These results are consistent with the mental model account of long-term memory.

Kole, J. A., Healy, A. F., & Bourne, L. E., Jr. (2008). Cognitive complications moderate the speed-accuracy tradeoff in data entry: A cognitive antidote to inhibition. *Applied Cognitive Psychology, 22,* 917-937.

Three experiments explored a speed-accuracy tradeoff reflecting decreasing response times (RTs) and increasing errors across trials in a data entry task. In Experiment 1, cognitive and motoric stressors were independently added to data entry, with the combination of stressors yielding the greatest decline in accuracy across blocks. Experiment 2 compared mental multiplication and simple data entry and manipulated the provision of feedback. Accuracy improved with both mental multiplication and feedback. Experiment 3 varied only the concluding keystroke; this extra requirement led to overall improvements in accuracy. In each experiment, RTs improved across trials. These results suggest that cognitive complications can serve as antidotes to inhibitory effects and can overcome the decline in accuracy due to continuous work on data entry.

Kole, J. A., Healy, A. F., Fierman, D. M., & Bourne, L. E., Jr. (2010). Contextual memory and skill transfer in category search. *Memory & Cognition, 38,* 67-82.

In three experiments, we examined transfer and contextual memory in a category search task. Each experiment included two phases (training and test), during which participants searched through category and exemplar menus for targets. In Experiment 1, the targets were from one of two domains during training (grocery store or department store); the domain was either the same or changed at test. Also, the categories were organized in one of two ways (alphabetically or semantically); the organization either remained the same or changed at test. In Experiments 2 and 3, domain and organization were held constant; however, categories or exemplars were the same, partially replaced, or entirely replaced across phases in order to simulate the dynamic nature of category search in everyday situations. Transfer occurred at test when the category organization or domain was maintained and when the categories or exemplars matched (partially or entirely) those at training. These results demonstrate that transfer is facilitated by overlap in training and testing contexts.

Krech Thomas, H., Healy, A. F., & Greenberg, S. N. (2007). Familiarization effects for bilingual letter detection involving translation or exact text repetition. *Canadian Journal of Experimental Psychology, 61*, 304-315.

In two experiments, English-Spanish bilinguals read passages, performing letter detection on some passages by circling target letters as they read. Detection passages were sometimes familiarized (primed) by prior reading of the same passage or a translation of it. Participants detected letters in English passages in Experiment 1 and in Spanish passages in Experiment 2. For both experiments, a missing letter effect occurred (depressed detection accuracy on frequent function words relative to less frequent content words). Familiarization promoted overall improvements in letter detection only for English passages, suggesting that reprocessing benefits depend on high language fluency. For Spanish passages, cognates engendered greater error rates than non-cognates; the visual similarity of Spanish and English cognates apparently enabled faster identification of Spanish cognates in a way unaffected by familiarization of the whole text passage. Priming by familiarized text was significantly higher when the passages were in the same language than when they were in different languages, suggesting that the reprocessing benefits are a with the GO model of reading (Greenberg, Healy, Koriat, & Kreiner, 2004) but require an expanded consideration of attention redistribution processes in that model.

Lohse, K. R., Sherwood, D. E., & Healy, A. F. (2010). How changing the focus of attention affects performance, kinematics, and electromyography in dart throwing. *Human Movement Science*, 29, 542-555.

Research has found an advantage for an external focus of attention in motor learning and control; instructing subjects to focus on the effects of their actions, rather than on body movements, can improve performance during training and retention testing. Previous research has mostly concentrated on movement outcomes, not on the quality of the movement itself. Thus, this study combined surface electromyography (EMG) with motion analysis and outcome measures in a dart throwing task, making this the first study that includes a comprehensive

analysis of changes in motor performance as a function of attentional focus. An external focus of attention led to better performance (less absolute error), decreased preparation time between throws, and reduced EMG activity in the triceps brachii. There was also some evidence of increased variability for kinematic measures of the shoulder joint under an external focus relative to an internal focus. These results suggest improved movement economy with an external focus of attention.

Overstreet, M. F., & Healy, A. F. (in press). Item and order information in semantic memory: Students' retention of the CU Fight Song lyrics. *Memory & Cognition*.

University of Colorado (CU) students were tested on memory for the *CU Fight Song* to examine serial position effects in semantic memory while controlling for familiarity across positions. In Experiment 1, students reconstructed the order of the 9 lines of the song. Students with previous exposure to the song performed better and showed a more bowed serial position function than students with no knowledge of the song. Experiment 2 added a task assessing memory of item information. One word was removed and replaced with a blank in each line, and an alternative word was offered as an option along with the correct word. Students selected the word that fit into each blank and then reconstructed the order of the lines. There was a bow-shaped curve for order reconstruction but not for item selection, which implies that the serial position function in semantic memory stems from order rather than item information.

Raymond, W. D., Healy, A. F., McDonnel, S., & Healy, C. A. (2009). Acquisition of morphological variation: The case of the English definite article. *Language and Cognitive Processes, 24,* 89-119.

Morphological systems have been pivotal in exploring cognitive mechanisms of language use and acquisition. Adult English definite article form preference seems to depend non-deterministically on multiple factors. A corpus study of adult spontaneous speech revealed similar patterns of variability. In an experiment, article variant preferences of three age groups were compared. Children were sensitive to the same phonological factors as adults, but showed effects of more limited experience with articulation and orthography. Preferences across age groups suggest developmental changes, but no evidence that children initially use a default form. Corpus studies of children's and adults' speech also revealed no evidence for a default. The results point to overgeneralisation of both article variants, resulting from extended competition between variant forms

Schneider, V. I., Healy, A. F., Barshi, I., & Kole, J. A. (in press). Following navigation instructions presented verbally or spatially: Effects on training, retention, and transfer. *Applied Cognitive Psychology*.

Two experiments investigated participants' ability to follow navigation instructions in a situation simulating communication between air traffic controllers and aircrews. A verbal condition, in which instructions were given

orally, was compared with a spatial condition, in which commands were shown on a computer display as simulated movements, with the presentation times in the two conditions equated. Retention and transfer were studied a week later when participants performed in either the same or the other condition. In both sessions, participants' initial proportion correct was much higher in the spatial than in the verbal condition, but after three blocks, accuracy in the two conditions was equivalent. Retention was perfect when training and test conditions matched. Training in the verbal condition transferred to the spatial condition but not vice versa. Thus, there is evidence that participants' representations of the movements in the verbal and spatial conditions were not equivalent.

Sumiya, H., & Healy, A. F. (2008). The Stroop effect in English-Japanese bilinguals: The effect of phonological similarity. *Experimental Psychology, 55,* 93-101.

English-Japanese bilinguals performed a Stroop color-word interference task with both English and Japanese stimuli and responded in both English and Japanese. The Japanese stimuli were either the traditional color terms (TCTs) written in Hiragana or loanwords (LWs) from English written in Katakana. Both within-language and between-language interference were found for all combinations of stimuli and responses. The between-language interference was larger for Katakana LWs (phonologically similar to English) than for Hiragana TCTs, especially with Japanese responses. The magnitude of this phonological effect increased with self-rated reading fluency in Japanese. Overall responding was slower and the Stroop effect larger with English than with Japanese stimuli. These results suggest that unintentional lexical access elicits automatic phonological processing even with intermediate-level reading proficiency.

Wohldmann, E. L., & Healy, A. F. (2010). Exploring specificity of speeded aiming movements: Examining different measures of transfer. *Memory & Cognition, 38,* 344-355.

Participants were trained and tested to move a mouse cursor from a start position to targets on a circular display in a perceptual–motor reversal condition, with horizontal, but not vertical, reversals. During training, some participants (experimental) moved to two targets either along a single diagonal axis (D1) or along both axes (D2). For D2, return movements from the targets were in the same direction as instructed movements to unpracticed targets. Others (control) trained on all targets. Testing always involved all targets. At test, movement times (to reach the target after leaving the start position) were shorter on trained than on untrained targets, especially for the D1 condition, documenting training specificity. However, movement times in the experimental conditions to new targets during testing were shorter than those in the control condition during training, documenting transfer of learning, with more transfer for D2 than for D1. Initiation times (to leave the start position after target onset) showed no transfer. The results provide evidence that specificity and transfer are not mutually exclusive and depend on the measure used to assess performance.

Wohldmann, E. L., Healy, A. F., & Bourne, L. E. Jr. (2007). Pushing the limits of imagination: Mental practice for learning sequences. *Journal of* Experimental *Psychology: Learning, Memory, and Cognition*, *33*, 254-261.

In 2 experiments, the efficacy of motor imagery for learning to type number sequences was examined. Adults practiced typing 4-digit numbers. Then, during subsequent training, they either typed in the same or a different location, imagined typing, merely looked at each number, or performed an irrelevant task. Repetition priming (faster responses for old relative to new numbers) was observed on an immediate test and after a 3-month delay for participants who imagined typing. Improvement across the delay in typing old and new numbers was found for the imagined and actual typing conditions but not for the other conditions. The findings suggest that imagery can be used to acquire and retain representations of sequences and to improve general typing skill.

Wohldmann, E. L., Healy, A. F., & Bourne, L. E., Jr. (2008). A mental practice superiority effect: Less retroactive interference and more transfer than physical practice. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *34*, 823-833.

Two experiments explored the benefits to retention and transfer conferred by mental practice. During familiarization, participants typed 4-digit numbers and took an immediate typing test on both old and new numbers. Participants then typed old 4-digit numbers, either physically or mentally, with either a different response configuration or the opposite hand from that used during familiarization. On a delayed test, participants physically typed both old and new numbers with the same response configuration and hand used during familiarization. Mental practice led to less retroactive interference and more transfer than did physical practice, supporting the hypothesis that mental practice strengthens an abstract representation that does not involve specific effectors.

Wohldmann, E. L., Healy, A. F., & Bourne, L. E., Jr. (2008). Global inhibition and midcourse corrections in speeded aiming. *Memory & Cognition, 36,* 1228-1235.

When some perceptual–motor relationships are reversed, participants might adopt a global inhibition strategy that replaces all normal movements with reversed movements. In two experiments, participants practiced moving a cursor from a start position to target locations. In a perceptual–motor reversal condition, in which horizontal but not vertical movements were reversed, participants were trained to move only to certain locations. Testing involved moving to all locations under the same reversal condition. Training on a subset of locations yielded partial transfer to untrained locations. These results support a global inhibition hypothesis modified to include both midcourse corrective movements and training specificity.

Wohldmann, E. L., Healy, A. F., & Bourne, L. E., Jr. (2010). Task integration in time production. *Attention, Perception, & Psychophysics*, *72,* 1130-1143.

Two experiments examined training on a prospective time production task. Participants produced intervals, expressed in fixed arbitrary units, while performing a concurrent secondary task. After a 15-min filled delay, the participants were retrained on the same tasks. These experiments tested whether the primary and secondary tasks would be integrated into a single task. In Experiment 1, the secondary task requirements were manipulated, but the time production task was fixed. In Experiment 2, the time production task requirements were manipulated, but the secondary task was fixed. The results suggest that participants integrate primary- and secondary-task requirements.

Young, M. D., Healy, A. F., Gonzalez, C., Dutt, V., & Bourne, L. E., Jr. (in press). Effects of training with added difficulties on RADAR detection. *Applied Cognitive Psychology*.

Three experiments simulating military RADAR detection addressed a training difficulty hypothesis (training with difficulty promotes superior later testing performance) and a procedural reinstatement hypothesis (test performance improves when training conditions match test conditions). Training and testing were separated by 1 week. Participants detected targets (either alphanumeric characters or vehicle pictures) occurring among distractors. Two secondary tasks were used to increase difficulty (a concurrent, irrelevant tone-counting task and a sequential, relevant action-firing response). In Experiment 1, involving alphanumeric targets with rapid displays, tone counting during training degraded test performance. In Experiment 2, involving vehicle targets with both sources of difficulty and slower presentation times, training under relevant difficulty aided test accuracy. In Experiment 3, involving vehicle targets and action firing with slow presentation times, test accuracy tended to be worst when neither training nor testing involved difficult conditions. These results show boundary conditions for the training difficulty and procedural reinstatement hypotheses.

Publications in Book Chapters or Conference Proceedings

Healy, A. F. (2007). Transfer: Specificity and generality. In H. L. Roediger, III, Y. Dudai, & S. M. Fitzpatrick (Eds.), *Science of memory: Concepts* (pp. 271-275). New York: Oxford University Press.

Healy, A. F., & Bonk, W. J. (2008). Serial learning. In H. L. Roediger, III (Ed.), *Cognitive* psychology *of memory* (pp. 53-63), Vol. 2 of *Learning and memory: A Comprehensive reference*, 4 vols. (J. Byrne, Editor). Oxford: Elsevier.

Healy, A. F., Kole, J. A., Wohldmann, E. L., Buck-Gengler, C. J., & Bourne, L. E., Jr. (in press). Data entry: A window to principles of training. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: A festschrift in honor of Robert A. Bjork*. New York: Psychology Press.

Studies reviewed are aimed to reveal principles of training, which lead to an understanding of what factors influence the efficiency, durability, and flexibility of training. The studies involve investigations of a simple data entry task. The

principles illustrated include principles derived from studies of word list learning – levels of processing and phonological processing – as well as newly formulated principles – procedural reinstatement, cognitive antidote, and mental practice. By the *depth of processing principle*, processing stimuli more deeply during training improves the skill involved in responding to those stimuli after a long delay. By the *phonological processing principle*, disrupting phonological processing of stimuli hinders the skill involved in responding to those stimuli but only when working memory is used to store the stimuli. By the *procedural reinstatement principle*, skill learning leads to durable retention when the required procedures are maintained but limited transfer when the required procedures are altered. By the *cognitive antidote principle*, adding cognitive complications to an otherwise routine task mitigates the adverse effects of prolonged work. By the *mental practice principle*, mental practice might have certain advantages over physical practice when it comes to slowing forgetting and promoting transfer of training because physical, but not mental, practice suffers from motoric interference when there is a change in effectors.

Healy, A. F., Schneider, V. I., & Barshi, I. (2009). Cognitive processes in communication between pilots and air traffic control. In E. B. Hartonek (Ed.), *Experimental psychology research trends* (pp. 45-77). Hauppauge, NY: Nova Science Publishers.

We have been probing the cognitive processes underlying communication between pilots and air traffic control. To study these processes, we developed an experimental paradigm analogous to the natural flight situation, in which pilots receive navigation instructions from air traffic control, repeat them, and follow them. In the experimental task, individuals typically hear navigation instructions, repeat them aloud, and then follow them, navigating in a space displayed on a computer screen. We describe a series of studies addressing 2 sets of relevant issues. The first set is empirical and concerns parameters for optimizing the ability to comprehend and remember the instructions, considering the length and wordiness of the instructions, the modality in which the instructions are presented, and the effects of repeating the instructions on their correct execution. The second set of issues is theoretical and concerns the mental representation of both the verbal content of the instructions and their spatial implications.

Healy, A. F., Wohldmann, E. L., Kole, J. A., Schneider, V. I., Shea, K. M., & Bourne, L. E., Jr. (in press). Training for efficient, durable, and flexible performance in the military. In W. Arthur, Jr., E. A. Day, W. Bennett, Jr., & A. Portrey (Eds.), *Individual and team skill decay: State of the science and implications for practice*. New York: Taylor & Francis.

Research is discussed on training, retention, and transfer of knowledge and skills. Optimal learning should be efficient, durable, and flexible. In the research discussed here, circumstances have been found leading to remarkable durability of what has been learned. Those same conditions, however, yield very poor flexibility, or the ability to generalize learning to new situations or contexts. A general theoretical framework is proposed that can account for the high degree

of specificity obtained in these studies but that also enables predictions of when learning will be generalizable rather than specific. The chapter is centered on five separate lines of research. The first three lines demonstrate a high degree of specificity of learning. These studies are presented by providing the empirical findings illustrating specificity and by briefly summarizing the theoretical explanations of them for the particular tasks investigated. The chapter ends with a summary of the results from the last two lines of research, which demonstrate, in support of the theoretical framework, situations showing robust transfer of learning. In summary, it is proposed that specificity (limited transfer) may occur for tasks based primarily on procedural information, or skill, whereas generalizability (robust transfer) may occur for tasks based primarily on declarative information, or facts. Thus, for skill learning, retention is strong but transfer is limited, whereas for fact learning, retention is poor but transfer is robust.

Staal, M. A., Bolton, A. E., Yaroush, R. A., & Bourne, L. E., Jr. (2008). Cognitive performance and resilience to stress. In B. Lukey & V. Tepe (Eds). *Biobehavioral resilience to stress* (pp. 259-300). London: Francis & Taylor.

Wickens, C. D., Ketels, S. L., Healy, A. F., Buck-Gengler, C. J., & Bourne, L. E., Jr. (in press). The anchoring heuristic in intelligence integration: A bias in need of de-biasing. *Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.

In information integration tasks, anchoring is a prominent heuristic, such that the first few arriving information sources (cues) tend to be given greater weight on the final integration product, than those cues following. Such a bias may be particularly problematic when the situation is dynamic, such that earlier arriving cues are more likely to have changed, and hence are less reliable for the final integration judgment. Such is often the case in military intelligence, when enemy intentions are inferred from multiple sources. We describe results of a simulation of such intelligence gathering in which anchoring is prominently manifest, in the processing of seven sequentially delivered cues bearing on enemy threat. In Experiment 1, an anchoring bias was present. In Experiment 2, a simple "de-biasing" wording inserted in the instructions and emphasizing the age of intelligence information induced more optimal weighting of the most recent cues,but did not eliminate anchoring.

Young, M. D., Wilson, M. L., & Healy, A. F. (2010). Improving reading skills for ESL learners using SoundSpel. In E. F. Caldwell (Ed.), *Bilinguals: Cognition, education and language processing* (215-227). Hauppauge, NY: Nova Science Publishers.

This study examined the effects of using a revised, transparent spelling system SoundSpel, a phonetic reading tool, with learners of English as a Second Language. During 6 training sessions, 12 participants used unaltered material and 12 used SoundSpel texts, in parallel with standard English, when reading American elementary school material. They then answered multiple-choice comprehension questions. Both groups were pre-tested and post-tested on

comprehension tests of similar elementary school material without SoundSpel. No group differences were found across tests or training (in quiz performance or reading time), suggesting no beneficial or harmful effects from using SoundSpel. A post hoc analysis suggested that SoundSpel would be most beneficial for students who learn to speak English before they learn to read it.

Manuscripts Submitted for Publication

Barshi, I., & Healy, A. F. (2010). *The effects of spatial representation on memory for verbal navigation instructions*. Manuscript submitted for publication.

Three experiments investigated effects of mental spatial representation on memory for verbal navigation instructions. The navigation instructions referred to a grid of stacked matrices displayed on a computer screen or on paper, with or without depth cues, and presented as two-dimensional diagrams or a three-dimensional physical model. Experimental instructions either did or did not promote a three-dimensional mental representation of the space. Subjects heard navigation instructions, immediately repeated them, and then followed them manually on the grid. In all display and experimental instruction conditions, memory for the navigation instructions was reduced when the task required mentally representing a three-dimensional space, with movements across multiple matrices, as compared with a two-dimensional space, with movements within a single matrix, even though the words in the navigation instructions were identical in all cases. The findings demonstrate that the mental representation of the space influences immediate verbatim memory for navigation instructions.

Healy, A. F., & Bourne, L. E., Jr. (2010). *Principles of training*. Manuscript submitted for publication.

The goal of our research has been to construct a theoretical and empirical framework that can account for and make accurate predictions about the effectiveness of different training methods for militarily relevant tasks. Towards this end, we have conducted basic research aimed to identify and empirically support training principles. We believe that the best way to transition our research to military applications is through these training principles. We trust that these principles can provide guidelines to trainers that will enhance the effectiveness of their training. We report four sets of experiments on the development and testing of training principles that illustrate the range of issues we have explored in our research. They include (a) tests of the generality across tasks of individual principles, (b) tests of multiple principles in a single task, (c) tests of principles in complex, dynamic environments, and (d) developing and testing new principles.

Healy, A. F., & Cunningham, T. (2010). *Detecting letters and words in prose passages*. Manuscript submitted for publication.

In 2 experiments, college students searched for either the letter *h* or the word *the* in prose passages in which every *h* occurred in the word *the*. In Experiment 1,

there were 3 passage versions, which differed only in that critical noun phrases were either *the* alone, "the word *the*," or "the definite article." More detection errors occurred for letter than for word targets, especially with "the definite article." In Experiment 2, there were 2 passage versions, which differed only in that a given noun phrase containing *the* occurred as a subject in one version and an object in the other. Again more detection errors occurred for letter than for letter sequence targets. Also, with letter targets but not with letter sequence targets, more detection errors occurred for object than for subject noun phrases. These findings suggest that both unitization and processing time contribute to detection errors in reading text.

Healy, A. F., & Greenberg, S. N. (2007). *Letter detection errors occur at two processing stages: Test of the GO Model*. Manuscript submitted for publication.

Students read prose passages and circled instances of the target letter *n* when the passages were printed normally or with 1-character-wide vertical stripes. More detection errors were made in the normal than in the striped condition. Detection errors were more frequent on the sequence –*ing* when it occurred as a word suffix than when it was embedded in a word stem. Passage format did not interact with word part (suffix, stem), and the effect of passage format was significant even in the striped condition, which hindered unitization processes. These results suggest that letter detection errors reflect processing occurring both during and after lexical access, in accordance with the GO model proposed by Greenberg, Healy, Koriat and Kreiner (2004).

Healy, A. F., Wohldmann, E. L., & Bourne, L. E., Jr. (2010). *Does practice with a defective mouse influence subsequent speeded aiming performance? A test of global inhibition*. Manuscript submitted for publication.

In a speeded aiming task, participants were trained to move a cursor with a mouse from a start position to target locations when the mouse-cursor relationships were either normal or defective (i.e., reversed vertically, horizontally, or both vertically and horizontally). Testing, which occurred after a 5-min delay, involved either the same or a different reversal condition. Response times improved across training, but no transfer occurred when reversal conditions were changed between training and testing. Specificity of training effects extended even to performance with the highly familiar normal mouse. Normal mouse use was slowed down by a factor of two to three with training on a defective mouse although the effect was transient in that case. Participants apparently adopt a global, rather than a local, inhibition strategy, suppressing all normal movements (and replacing them with sensorimotor remapped movements) but disinhibiting movements along any non-reversed dimension (selectively disengaging the sensorimotor remapping).

Kole, J. A., & Healy, A. F. (2010). *Memory for details about people: Familiarity, relatedness, and gender congruency*. Manuscript submitted for publication.

Several recent studies have demonstrated that processing information in terms of survival value improves retention over short delays. These findings are

interpreted within a functionalist framework, which posits that modern cognitive processes reflect ancient selection pressures. The present study examines factors that influence memory for details about people. In 2 experiments, subjects learned fictitious details about familiar (friends, relatives) and/or unfamiliar individuals, and were tested both immediately and after a 1-week delay. To control for a confounding between familiarity and genetic relatedness in Experiment 1, in Experiment 2 specific relationships (identical twin, first cousin, acquaintance) were assigned to unfamiliar individuals. Across experiments, retention was enhanced for familiar compared to unfamiliar individuals, for friends/acquaintances compared to relatives, for more closely than distantly related individuals, and for individuals of the opposite gender as the subject. The results are consistent with a functionalist framework when both mate selection and kin are considered.

Krech Thomas, H., & Healy, A. F. (2010). *A comparison of rereading benefits in first- and second-language reading*. Manuscript submitted for publication.

Text comprehension models in first and second language reading research posit that slow word recognition inhibits reading speed and decreases comprehension. To investigate the role of word recognition in reading, 2 experiments examined rereading benefits in participants' first and second languages using scrambled and normal versions of English and Spanish texts. Native English speakers with intermediate (Experiment 1) or advanced (Experiment 2) Spanish skills demonstrated word- and text-level transfer in both English and Spanish. However, advanced Spanish readers did not exhibit word-level transfer when reading simple Spanish texts. These results suggest that fluent reading may be strongly influenced not just by word recognition, but also by text difficulty relative to reader skill, as well as other factors.

LaVoie, N. N., Healy, A. F., & Bourne, L. E., Jr. (2006). *Seeding in a qualitative domain: Sound-spelling mappings in French*. Manuscript submitted for publication.

Two experiments examined the acquisition of sound-spelling mappings in an unfamiliar language (French) using a controlled seeding paradigm (LaVoie, Bourne, & Healy, 2002). Participated were required to spell lists of spoken French words in a pretest, seeding phase, posttest, and 2-week retention test. The words in each list varied in the size of their phonological neighborhoods and in their frequencies of occurrence. Spelling was better on the posttest than on the pretest, and this seeding effect was maintained across the retention interval, despite some forgetting of individual seed words. Effects of neighborhood size and frequency were minor and evident only at the retention test, suggesting that a single example is sufficient to seed the sound-spelling knowledge base.

Lohse, K. R., & Healy, A. F. (2009). *Exploring the contributions of declarative and procedural information to training: A test of the procedural reinstatement principle*. Manuscript submitted for publication.

According to the procedural reinstatement principle, procedural training leads to

strong retention but limited transfer, whereas declarative training leads to poor retention but robust transfer. To test this principle in Experiment 1, participants were trained in one of 3 conditions (declarative, procedural, mixed) and were subsequently tested in either the same or another condition. The task and required responses were the same in the three conditions; they differed only in the emphasis given to declarative or procedural information. Consistent with the procedural reinstatement principle, in terms of response time procedural training was more durable than declarative training. In Experiment 2, transfer was assessed using procedural and declarative conditions, but participants transferred between tasks within those conditions. Although there was transfer in response time between tasks with procedural training, the greatest magnitude of transfer was found in one direction with declarative training, again consistent with the procedural reinstatement principle.

Lohse, K. R., Healy, A. F., & Sherwood, D. E. (2010). *Mental practice in the intermanual transfer of motor skills*. Manuscript submitted for publication.

The current study compared intermanual transfer for two different handwriting tasks (familiar letters and novel symbols), following both mental and physical practice. There was substantial transfer from practice with the dominant to the nondominant hand in both time to produce a character and size of the character produced, but no transfer in the reverse direction (even for novel symbols). Most importantly, there was significant transfer as a result of mental practice in production time comparable to physical practice. However, there was no transfer from mental practice when measuring character size. During mental practice, task-level variables still had significant effects, whereas effector-level variables did not. Thus, asymmetrical transfer as a result of mental practice is posited to result from the transfer of task-level processes but not effector-level processes.

Lohse, K. R., Sherwood, D. E., & Healy, A. F. (2010). *Neuromuscular effects of shifting the focus of attention in a simple force production task*. Manuscript submitted for publication.

Research on the focus of attention has begun exploring the physiological changes that underlie the shift between internal and external foci of attention. However, previous electromyography studies have used dynamic tasks, making it difficult to interpret electrophysiological data. The current experiment analyzed how the focus of attention affects a subject's ability to perform an isometric force production task (focus was directed either at the force platform or the muscles responsible for force production). Subjects received training without attentional focus instructions and then completed blocks of trials with either an internal or external attentional focus. An external focus led to significantly less error, reduced EMG activity, and lower median power frequencies in the antagonist muscle, but had no effect on the agonist muscle. Thus, an external focus of attention led to more efficient motor unit recruitment patterns (reduced cocontraction) and improved performance. Post-test surveys revealed subjects' were aware of their improved performance with an external focus.

Raymond, W. D., Healy, A. F., & McDonnel, S. J. (2010). *Pairing words with syntactic frames: Syntax, semantics, and count-mass usage*. Manuscript submitted for publication.

Two experiments examined English speakers' choices of count or mass compatible frames for nouns varying in imageability (concrete, abstract) and noun class (count, mass). Pairing preferences with equative (*much/many*) and non-equative (*less/fewer*) constructions were compared for groups of teenagers, young adults, and older adults. Deviations from normative usage were, for all ages, larger for count than for mass nouns, for the non-equative than for the equative construction, and for abstract count and concrete mass words than for the other combinations. These results indicate that mass syntax is not a developmental default, support proposals that mass syntax is more flexible than count syntax, verify the non-prescriptive use of *less* with count nouns, and extend the interaction of syntax and semantics in noun classification to older ages, with older adults showing a reduced reliance on semantics. Knowledge of frame compatibility and knowledge of noun class are also shown to be largely independent.

Wilson, M. L., & Healy, A. F. (2008). *Effects of time pressure on mood and performance*. Manuscript submitted for publication.

This 2-part study explored the effects of time pressure on mood and cognition using self-reports of mood and behavioral measurements of cognitive performance. In both parts, participants filled out a Positive and Negative Affect Schedule (PANAS) form before and after taking brief spatial, verbal, and math intelligent quotient (IQ) tests. Of the 144 participants included in the analyses, 72 experienced time pressure in Part 1, and 72 experienced time pressure in Part 2—with the 2 parts separated by 1 week. Test condition influenced negative (but not positive) affect and performance under all three IQ tests. Under time pressure, negative affect increased, and the number of correct IQ test responses declined. Results suggest that time pressure can depress mood and hinder cognitive performance. These findings have important implications for assessment of academic ability.

Papers Presented at Meetings

Anderson, L. S., Healy, A. F., Kole, J. A., & Bourne, L. E., Jr. (2010, April). *The clicker technique: An effective method of teaching compression*. Poster presented at the 80th Annual Convention of the Rocky Mountain Psychological Association, Denver, CO.

Astbury, B. Schneider, V. I., Healy, A. F., Barshi, I., & Bourne, L. E., Jr. (2008, April). *The effects of blocking or mixing message length in a navigational paradigm*. Poster presented at the 78th Annual Convention of the Rocky Mountain Psychological Association, Boise, ID.

Bonk, B. (2006, April). *Sequence memory with visual item, spatial, and order information*. Paper presented at the 76th annual convention of the Rocky Mountain *Psychological* Association, Park City, Utah.

Bourne, L. E., Jr., Healy, A. F., Bonk, W. J., & Buck-Gengler, C. J. (2009, November). *Prospective memory offers some protection against forgetting associated items*. Paper presented at the 50th Annual Meeting of the Psychonomic Society, Boston, MA.

Buck-Gengler, C. J., Bonk, W. J., Healy, A. F., & Bourne, L. E., Jr. (2007, April). *The memory constriction hypothesis: Retrospective and prospective memory under time pressure*. Paper presented at the 77th Annual Convention of the Rocky Mountain Psychological Association Meeting, Denver, CO.

Cunningham, T. F., & Healy, A. F. (2006, November). *Is word detection better than letter detection while reading prose?* Poster presented at the 47th Annual Meeting of the Psychonomic Society, Houston, TX.

Cunningham, T. F., Healy, A. F., Shea, K. M., & Kole, J. A. (2007, July). *Distinctiveness, but not familiarity, affects reconstruction of order in immediate episodic memory*. Poster presented at the joint meeting of the Experimental Psychology Society and the Psychonomic Society, Edinburgh, Scotland.

Fierman, D. M., & Healy, A. F. (2007, April). *Unintentional translation in the sentence-level bilingual Stroop task*. Paper presented at the 77th Annual Convention of the Rocky Mountain Psychological Association Meeting, Denver, CO.

Fierman, D. M., Healy, A. F., & Bourne, L. E., Jr. (2007, August). *Optimizing memory for instructions by varying presentation modality: Explorations of a navigation task*. Invited poster presented in the Symposium on Memory Dynamics and the Optimization of Instruction. American Psychological Association, San Francisco, CA.

Healy, A. F. (2006, August). *What we know and what we need to know in learning science to achieve greater efficiency and effectiveness in training*. Invited paper presented at the Army Science of Learning Workshop. Hampton, VA.

Healy, A. F. (2007, May). *Training, retention, and transfer of knowledge and skills*. Invited talk delivered at the 11th International Conference on Cognitive and Neural Systems. Boston University, Boston, MA.

Healy, A. F. (2008, April). *Training optimization: Learning, retention, and transfer of knowledge and skills*. Paper presented at the Meeting of the Society of Experimental Psychologists, Bloomington, IN.

Healy, A. F. (2009, January). *Data entry: A window to principles of training*. Invited paper presented at the conference "Successful remembering and successful forgetting: A Festschrift in honor of Robert A. Bjork," Los Angeles, CA.

Healy, A. F. (2009, May). *Data entry: A window to principles of training*. Paper presented at the 106th Annual Meeting of the Society of Experimental Psychologists, Boulder, CO.

Healy, A. F. (2009, October). *Principles of training*. Invited paper presented at the Workshop to Explore Issues and Mitigation Strategies for Long Term Retention of Military Expertise. Mesa, Arizona.

Healy, A. F. (2010, April). *Experiments on development and testing of training principles*. Paper presented at the 2010 Ellis-Battig Memory Symposium: Optimizing the training of knowledge and skills: A review of accomplishments from the Multidisciplinary University Research Initiative (MURI) on training, 80th Annual Convention of the Rocky Mountain Psychological Association, Denver, CO.

Healy, A. F., & Cunningham, T. F. (2009, November). *Detection of letter and letter sequence targets while processing prose*. Poster presented at the 50th Annual Meeting of the Psychonomic Society, Boston, MA.

Healy, A. F. Wohldmann, E. L., & Bourne, L. E., Jr. (2007, July). *Training specificity and global inhibition in speeded aiming movements*. Paper presented at the Joint Meeting of the Psychonomic Society and the UK Experimental Psychology Society, Edinburgh, Scotland.

Jones, M., Worthy, D. A., Ketels, S. L., & Otto, A. R. (2010, August). *The phenomenology of multiple learning systems*. Paper to be presented at the Ninth Annual Summer Interdisciplinary Conference, Bend, OR.

Ketels, S. (2008, July). *Implicit category sequence learning*. Poster presented at the XXIX International Congress of Psychology, Berlin, Germany.

Ketels, S. L., Healy, A. F., Wickens, C. D., Buck-Gengler, C. J., & Bourne, L. E., Jr. (2010, April). *Spatial list learning and decision making in the fusion paradigm*. Poster presented at the 80th Annual Convention of the Rocky Mountain Psychological Association, Denver, CO.

Kole, J. A., & Healy, A. F. (2007, November). *Examining the retrieval process in mediated learning using priming effects*. Poster presented at the 48th Annual Meeting of the Psychonomic Society, Long Beach, CA.

Kole, J. A., & Healy, A. F. (2008, November). *Retrieval-induced forgetting: Examining representation weakening and retrieval competition accounts*. Poster presented at the 49th Annual Meeting of the Psychonomic Society, Chicago, IL.

Kole, J. A., & Healy, A. F. (2009, November). *Long-term retention of knowledge about friends, family, and unfamiliar individuals*. Poster presented at the 50th Annual Meeting of the Psychonomic Society, Boston, MA.

Kole, J. A., & Healy, A. F. (2010, April). *Memory for facts about people: Familiarity, relatedness, degree of genetic similarity, and gender congruency*. Poster presented at the 80th Annual Convention of the Rocky Mountain Psychological Association, Denver, CO.

Kole, J. A., Healy, A. F., & Bourne, L. E., Jr. (2006, November). *A cognitive antidote to inhibition: Data entry under conditions of prolonged work.* Poster presented at the 47th Annual Meeting of the Psychonomic Society, Houston, TX.

Lohse, K. R. (2008, April). *Intermanual transfer in handwriting: Evidence of asymmetrical transfer and priming.* Paper presented at the 78th Annual *Convention* of the Rocky Mountain Psychological Association Meeting, Boise, ID.

Lohse, K. R., & Healy, A. F. (2008, November). *Exploring the contributions of declarative and procedural training to performance*. Poster presented at the 49th Annual Meeting of the Psychonomic Society, Chicago, IL.

Lohse, K. R., Healy, A. F., & Sherwood, D. E. (2009, November). *Task-level and effector-level representations in intermanual transfer of motor skills*. Poster presented at the 50th Annual Meeting of the Psychonomic Society, Boston, MA.

Lohse, K. R., Sherwood, D. E., & Healy, A. F. (2010, April). *How changing the focus of attention affects performance, kinematics, and electromyography in dart throwing*. Paper presented at the 80th Annual Convention of the Rocky Mountain Psychological Association, Denver, CO.

Lohse, K. R., Sherwood, D. E., & Healy, A. F. (2010, June). *How changing the focus of attention affects performance, kinematics, and electromyography*. Paper presented at NASPSA (North American Society for Psychology of Sport and Physical Activity) Conference 2010, Tucson, AZ.

Lohse, K. R., Sherwood, D. E., & Healy, A. F. (2010, June). *The influence of attention on learning and performance: Two experiments in isometric force production*. Poster presented at NASPSA (North American Society for Psychology of Sport and Physical Activity) Conference 2010, Tucson, AZ.

McCormick, B., & Healy, A. F. (2010, April). *Words and symbols use different working memory resources in a navigational task*. Poster presented at the 80th Annual Convention of the Rocky Mountain Psychological Association, Denver, CO.

Overstreet, M. F., & Healy, A. F. (2010, April). *Item and order information in semantic memory: Students' retention of the CU fight song*. Poster presented at the 80th Annual Convention of the Rocky Mountain Psychological Association, Denver, CO.

Raymond, W. D., Buck-Gengler, C. J., Healy, A. F., & Bourne, L. E., Jr. (2007, April). *Predicting differences in individual learning and performance in a motor skill task*. Paper presented at the 77th Annual Convention of the Rocky Mountain Psychological Association Meeting, Denver, CO.

Raymond, W. D., Healy, A. F., & McDonnel, S. (2009, April). *Lexical, syntactic, and semantic influences on count-mass preferences by teenagers and adults*. Poster presented at the Biennial Meeting of the Society for Research in Child Development. Denver, CO.

Raymond, W. D., Healy, A. F., Rains, J., & Bourne, L. E., Jr. (2006, April). *Affordances and choices in strategy shifts in skill acquisition tasks*. Paper presented at the 76th annual convention of the Rocky Mountain Psychological Association, Park City, Utah.

Schneider, V. I., Healy, A. F., Barshi, I., & Kole, J. A. (2006, November). *Following verbal and spatial navigation instructions: Training, retention, and transfer.* Paper presented at the 47th Annual Meeting of the Psychonomic Society, Houston, TX.

Schneider, V. I., Healy, A. F., Barshi, I., McCormick, B., & Bourne, L. E., Jr. (2009, November). *Effects of presentation order during training to follow navigation instructions*. Poster presented at the 50th Annual Meeting of the Psychonomic Society, Boston, MA.

Schneider, V. I., Healy, A. F., Barshi, I., & Bourne, L. E., Jr. (2007, November). *Effects of difficulty, specificity, and variability on training to follow navigation instructions*. Poster presented at the 48th Annual Meeting of the Psychonomic Society, Long Beach, CA.

Schneider, V. I., Healy, A. F., Barshi, I, & Parker, J. T. (2005, November). *Effect of computer display on executing navigation instructions*. Poster presented at the 46th Annual Meeting of the Psychonomic Society, Toronto, Canada.

Schneider, V. I., Healy, A. F., Buck-Gengler, C. J., Barshi, I., & Bourne, L. E., Jr. (2007, July). *The effects of feedback on learning to follow navigation instructions*. Poster presented at the joint meeting of the Experimental Psychology Society and the Psychonomic Society, Edinburgh, Scotland.

Schneider, V. I., Healy, A. F., Buck-Gengler, C. J., Barshi, I., & Bourne, L. E., Jr. (2008, November). *Effects of presenting navigation instructions twice in the same or different modalities*. Paper presented at the 49th Annual Meeting of the Psychonomic Society, Chicago, IL.

Tao, L., & Healy, A. F. (2006, November). *Stroop effect in Chinese characters and Pinyin: Orthography and language experience*. Poster presented at the 47th Annual Meeting of the Psychonomic Society, Houston, TX.

Tao, L., & Healy, A. F. (2007, November). *Unitization effect in English and Chinese: Orthography and language experience*. Poster presented at the 48th Annual Meeting of the Psychonomic Society, Long Beach, CA.

Wohldmann, E. L., Healy, A. F., & Bourne, L.E., Jr. (2005, November). *Imagine that! Motor imagery enhances repetition priming of sequences*. Poster presented at the 46th Annual Meeting of the Psychonomic Society, Toronto, Canada.

Wohldmann, E. L., Healy, A. F., & Bourne, L. E., Jr. (2006, April). *Use your imagination: Learning and maintaining representations in a sequential motor task.* Paper presented at the 76th annual convention of the Rocky Mountain Psychological Association, Park City, Utah.

Wohldmann, E. L., Healy, A. F., & & Bourne, L. E., Jr. (2006, November). *Mental practice leads to less forgetting and interference than physical practice*. Paper presented *at* the 47th Annual Meeting of the Psychonomic Society, Houston, TX.

Wohldmann, E. L., Healy, A. F., & Bourne, L. E., Jr. (2007, November). *Physical but not mental practice yields retroactive interference*. Paper presented at the 48th Annual Meeting of the Psychonomic Society, Long Beach, CA.

Wohldmann, E. L., & Healy, A. F. (2008, November). *Perceptual and motoric specificity in training of speeded aiming movements*. Paper presented at the 49th Annual Meeting of the Psychonomic Society, Chicago, IL.

Young, M. D., Healy, A. F., & Bourne, L. E., Jr. (2009, November). *Training and transfer of an artificial grammar*. Poster presented at the 50th Annual Meeting of the Psychonomic Society, Boston, MA.

Young, M. D., Healy, A. F., & Bourne, L. E., Jr. (2010, April). *Artificial grammar learning: Retention and transfer*. Poster presented at the 80th Annual Convention of the Rocky Mountain Psychological Association, Denver, CO.

Young, M. D., Healy, A. F., Gonzalez, C. & Bourne, L. E., Jr. (2007, July). *The effects of training difficulty on RADAR detection*. Poster presented at the joint *meeting* of the Experimental Psychology Society and the Psychonomic Society, Edinburgh, Scotland.

Young, M. D., Healy, A. F., Gonzalez, C., Dutt, V., & Bourne, L. E., Jr. (2008, November). *Effects of training with added relevant responses on RADAR detection*. Poster presented at the 49th Annual Meeting of the Psychonomic Society, Chicago, IL.

Young, M. D., Wilson, M. L, & Healy, A. F. (2008, April). *Improving reading skills For ESL learners using SoundSpel*. Poster presented at the 78th Annual Convention of the Rocky Mountain Psychological Association Meeting, Boise, ID.

Doctoral Dissertation and Master's Theses

Bonk, W. J. (2008). *Serial learning of verbal and nonverbal sequences*. Unpublished doctoral dissertation, University of Colorado, Boulder.

Ketels, S. (2009). *Implicit category sequence learning*. Unpublished masters thesis, University of Colorado, Boulder.

Kole, J. A. (2007). *The retrieval process in mediated learning: Using priming effects to test the direct access and covert mediation models*. Unpublished doctoral dissertation, University of Colorado, Boulder.

Lohse, K. R. (2009). *Task level and effector level representations in the intermanual transfer of* fine *motor skills: The effects of task familiarity and mental practice*. Unpublished masters thesis, University of Colorado, Boulder.

Wohldmann, E. L. (2006). *Pushing the limits of imagination: The effectiveness of motor imagery for acquiring and maintaining a sequential motor skill*. Unpublished doctoral dissertation, University of Colorado, Boulder.

Young, M. D. (2010). *Artificial grammar learning: Implicit and explicit components for retention and transfer*. Unpublished doctoral dissertation, University of Colorado, Boulder.

Towards the Improvement of Astronaut Training: A Literature Review of Empirical Evidence for Training Principles

Alice F. Healy, Vivian I. Schneider, and Lyle E. Bourne, Jr.
University of Colorado, Boulder

Towards the Improvement of Astronaut Training:
A Literature Review of Empirical Evidence for Training Principles

Alice F. Healy, Vivian I. Schneider, and Lyle E. Bourne, Jr.
University of Colorado, Boulder

I. Introduction

A. Purpose of this review

This document reviews the existing literature on theoretical and empirical research in experimental cognitive psychology as it pertains to training, with a particular focus on the training of astronauts and other military personnel. The aim is to identify evidence-based principles of training that are well enough established that they might be implemented in actual training regimens. The principles vary to some degree in their empirical support, but this review includes only those for which there is convincing evidence and theoretical understanding. Nevertheless, for purposes of organization, those principles that are strongly established are distinguished from those that are promising but require additional validation.

B. Some important distinctions

There are some important distinctions to keep in mind that influence the organization of this document and the implications that can be drawn from it.

1. Training principles, guidelines, and specifications

The most important distinction is one raised by Salas, Cannon-Bowers, and Blickensderfer (1999) among training principles, training guidelines, and training specifications. Principles, guidelines, and specifications all relate to how training is best accomplished. In effect, they provide a conduit between training theory and training practice. A principle, which is the level addressed in this review, is an underlying truth or fact about human behavior. A guideline, in contrast, is a description of actions or conditions that, if correctly applied, could improve training. A specification is a detailed, precise statement of how training should be designed by operationalizing training guidelines in the development of training programs. This review, thus, provides an initial step towards designing training programs that can optimize on-the-job performance. Additional developmental or applied research will be required to translate these principles into guidelines and, subsequently, to specifications. This review focuses primarily on training principles but also offers suggested guidelines that might be examined in further research.

2. Training vs. education

People generally think of training and education as being essentially the same. However, in this paper, a distinction is drawn between these processes. Education relates

to general knowledge and skills identified with particular domains, such as history or physics. Training, in contrast, relates to particular jobs or tasks that also require knowledge and skills but are more specific to the goals of those activities. Thus, principles of training are tied to the improvement of performance of duties in particular occupations, such as electrician or computer programmer. The principles of training are not necessarily the same as principles of education although there is undoubtedly a good deal of overlap. Both training and education represent a transaction between teachers and students. The principles of training considered here recognize that relationship and apply to both teachers and students.

3. Training of knowledge vs. training of skills

The principles discussed here apply to both declarative information (knowledge) and procedural information (skills). Knowledge consists of facts, discriminations, and concepts about a domain, which are generally explicit and a part of a person's awareness about a given task. In contrast, skills consist of knowing how to use those facts, which might be implicit and outside of a person's awareness or consciousness. For example, in statistics, knowledge includes the fact that the standard deviation is a measure of data dispersion, whereas skills include executing the sequence of steps needed to compute a standard deviation in a data set. Both knowledge and skills are hierarchical and are logically linked together; facts at every level of abstraction are associated with procedures for using them. Note that training applies primarily to skill learning, whereas education emphasizes fact learning, although fact and skill learning are involved in both training and education.

C. Scope of this review

Principles of training will be reviewed for which there is at least some experimental evidence. The principles will be presented in categories or clusters. One basis of this organization is the degree of empirical support because some principles are strongly supported by the evidence, whereas the evidence for others is partial and incomplete. Within these broad categories, grouping relies on similarity of effects. It should be recognized at the outset that both these broad and more specific categories are somewhat arbitrary. A given principle might have been categorized differently or placed in more than one category, but only a single category choice was used here. Where necessary, cross linkages between categories are referenced.

II. Fundamental cognitive processes underlying training

Training implicates three fundamental underlying cognitive processes: acquisition (learning), retention (memory), and transfer (generalization). There are basic principles that apply at the level of these fundamental processes, which are the starting point of the review.

A. Acquisition: Power law of practice

There are two major measures of performance during the acquisition of knowledge and skills: accuracy and speed of responses. With respect to response speed, Newell and Rosenbloom (1981) have argued that the *Power Law of Practice* describes the acquisition process for most skills. This law formalizes the relationship between trials of practice and time to make a correct response as a power function, $R = aN^{-b}$, where R is response time on trial N, a is response time on trial 1, and b is the rate of change. It follows that the relationship between response time and trial number is linear in log-log coordinates, log R = log a – b log N. In some cases, where more than one strategy can be used in the task, separate power functions apply to the different strategies (Delaney, Reder, Staszewski, & Ritter, 1998; Rickard, 1997). This principle affords a way of predicting performance in a variety of tasks as a function of degree of practice (but see Roediger, 2008). With respect to response accuracy, a similar function seems to apply (e.g., Bourne, Healy, Parker, & Rickard, 1999) although a power function has not been proposed for such data.

In some cases, speed and accuracy might not be positively correlated (e.g., Pachella, 1974). People sometimes trade speed for accuracy or vice versa. Likewise, the speed of executing the different steps of a complex task may not be positively correlated, with people slowing down on one step in order to be faster on another step (Healy, Kole, Buck-Gengler, & Bourne, 2004; Kole, Healy, & Bourne, 2008). In these cases, the power law of practice might not be a good description for all measures. Furthermore, for optimal training, instructors need to be aware of what are the various steps in any task as well as whether speed or accuracy is more important in each step, so that the more important aspect can be emphasized in training.

B. Retention: Power law of forgetting

With the passage of time and the lack of opportunity to rehearse or refresh acquired knowledge or skills, performance declines, reflecting forgetting of what was learned. This decline in performance, exhibited in increased response time (or decreased accuracy), has been known since the time of Ebbinghaus (1885/1913), who used a measure of savings (i.e., the amount of relearning required to achieve the criterion level of performance during original learning). Subsequently this relationship between response time and retention interval was described as a power law (Wickelgren, 1974), $R = d + fT^{-g}$, where R is response time, T is the retention interval, d is the criterion of original learning, f is a scaling parameter, and g is the rate of forgetting. This *Power Law of Forgetting* (Wixted & Carpenter, 2007; see also Rubin & Wenzel, 1996) can be thought of as the inverse of the power law of practice (but see Roediger, 2008).

C. Transfer: Laws relating to similarity

Training on a particular task has implications for performance on other related tasks. The effect of training on one task can be either positive (facilitation) or negative (interference) on performance of another task. When the acquisition of one task affects performance on another, that effect is called transfer. The major variable determining the

extent and direction of transfer is similarity between the two tasks. Osgood (1949) has conceptualized this relationship in the form of a transfer surface, which relates transfer magnitude both to response similarity and to stimulus similarity between the training and the transfer tasks. When the stimuli in the two tasks are varied in their similarity and the responses are held constant, positive transfer is obtained, with its magnitude increasing as the similarity between the stimuli increases. On the other hand, when the stimuli are held constant and the responses are varied in their similarity, negative transfer is obtained, with its magnitude decreasing as the similarity between the responses increases. Finally, when both the stimuli and responses are simultaneously varied in their similarity, negative transfer is obtained, with its magnitude increasing as the similarity between stimuli increases. Shepard (1987) has given a quantitative expression to such similarity functions, which he refers to as a universal law of generalization.

III. Well established training principles

Well established training principles will now be reviewed, under the following categories: (a) resource and effort allocation, (b) context effects, (c) task parameters, and (d) individual differences. Again, readers should keep in mind that the category scheme is arbitrary and that a given principle might be relevant to more than one category.

A. Principles relating to resource and effort allocation

Implementation of some training principles requires the learner to direct or allocate cognitive resources and effort to particular aspects of the knowledge or skills to be acquired.

1. Deliberate practice

Practice makes perfect, but not all practice is equivalent in terms of its effectiveness. Deliberate (i.e., highly focused and highly motivated) practice is best in terms of promoting skill acquisition and expertise (Ericsson, Krampe, & Tesch-Römer, 1993). Indeed, learners, even those who might be highly talented or have a high aptitude for the training domain, will not acquire their highest level of performance if they do not engage in deliberate practice over a prolonged period of time with many repetitions of the skill to be performed. *Guideline*: By initial instructions to trainees, try to engage deliberate practice at the outset and throughout the training process.

2. Depth of processing

One aspect of deliberate practice relates to how deeply the material to be learned is processed. Activities during training that promote deep or elaborate processing of materials yield superior retention (e.g., Craik & Lockhart, 1972; but see Roediger, 2008). The depth of processing principle can be achieved in various ways, including simply presenting the material in a format that requires a translation process or speech coding. Counter to intuition, when numerical data must be entered into some system, the numbers should be presented in word format (e.g., three-five-two) rather than numeral format (3-

5-2) to maximize memory for the numbers. Word format, but not numeral format, requires translation from the words to the digits represented on a keyboard and facilitates speech coding of the digits. This additional process enhances long-term memory for the material (Buck-Gengler & Healy, 2001). *Guideline*: To enhance the durability of training material, promote deep processing of the material to be learned either by explicit instructions or by incidental task demands.

3. Active versus passive learning

In general, it is better to use active learning rather than passive learning techniques. For example, if the task is to memorize a set of procedures for troubleshooting a piece of equipment, the trainees should try to generate the procedures from memory, rather than simply to read or reread them. Then the trainees should check the accuracy of their actively generated responses against the correct list and make note of any errors. They should actively generate the list again until they are able to produce it without error. This recommendation follows from the generation effect (the finding that people show better retention of learned material when it is self-produced, or generated, than when it is simply copied or read; e.g., Crutcher & Healy, 1989; McNamara & Healy, 1995, 2000; Slamecka & Graf, 1978; but see Roediger, 2008).

More generally, a trainee is typically passive, with the trainer controlling the course of events during training. However, there is evidence to believe that actively involving the trainee in the learning process facilitates training efficiency and the level of achievement reached (see, e.g., Hockey & Earle, 2006; Norman, 2004; Péruch & Wilson, 2004; Vakil, Hoffman, & Myzliek, 1998). Active involvement entails some self-regulation by the trainee. There has been relatively little research focused, however, on the self-regulation process and on the self-regulation skill (Perels, Gürtler, & Schmitz, 2005; Schunk, 2005). There are, though, some basic cognitive processes related to active learning and self-regulation that have been studied in detail. Among those processes are the aforementioned generation effect, metacognition (e.g., Mazzoni & Nelson, 1998; Sperling, Howard, Staley, & DuBois, 2004), and discovery learning (e.g., McDaniel & Schlager, 1990). It is possible that self-regulation might enhance training efficiency, and it is also possible that self-regulation might have a positive impact on the durability of skills and their transfer to performance in new contexts although there is little relevant evidence presently available.

Bjork, deWinstanley, and Storm (2007) make three points about learners that are relevant to self-regulation: (a) Learners often are quite inaccurate when monitoring their level of comprehension about material they are studying. (b) How learners rate their comprehension determines how they allocate resources for further study, allocating more resources to those aspects of the material that they do not yet understand. (c) Learners can inaccurately assess their comprehension because of “illusions of comprehension,” which are caused by specific learning methods, such as massed practice, which might lead to good performance during study but to poor long-term retention or transfer (Bjork, 1999; Simon & Bjork, 2001).

Bjork et al. (2007) examined whether or not students can discover the benefits of using generation for learning and then put it into use as they study (deWinstanley & Bjork, 2004; Koriat, 1997). Making students aware of the benefits of generation as a learning tool led them to adopt better strategies for encoding new information while studying. However, just putting students in a condition that requires generation is not likely to induce students to discover and then adopt the more effective strategies in subsequent study times. Students might need to experience the results of different study methods before they can appreciate which methods are more effective. These self-identified methods can then be used for later learning and study activities.

Kornell and Bjork (2007) found that students make study decisions by what is more urgent at the moment (usually last minute cramming) rather than by trying to maximize long-term learning. Students need to learn how to learn (Bjork, 2001). They conclude that for students to enhance their long-term memory they need to know how learning works and use that knowledge to go against some of their intuitions and indices of short-term memory.

Guideline: Trainers should use whatever methods are possible to engage trainees actively in the learning process, including requiring them to generate answers to questions periodically, instructing them directly or indirectly to maintain awareness about their progress in learning, and allowing them to experience the consequences of their study strategy.

B. Principles relating to context effects

Some training principles reflect the fact that training is often context specific, meaning that the knowledge and skills learned are bound, at least to some degree, to the circumstances in which they were acquired. The following are the two most important, well-established principles of this type.

1. Procedural reinstatement

The procedural reinstatement principle implies that duplicating at test procedures that were required during learning facilitates subsequent retention and transfer (Clawson, Healy, Ericsson, & Bourne, 2001; Healy et al., 1992; Healy, Wohldmann, & Bourne, 2005). This principle is similar to others that had been derived primarily from studies of list learning, including the principles of encoding specificity (memory for information is best when retrieval cues elicit the original encoding operations; e.g., Tulving & Thomson, 1973), transfer appropriate processing (memory performance will be best when test procedures evoke the procedures used during prior learning; e.g., Morris, Bransford, & Franks, 1977; Roediger, Weldon, & Challis, 1989), and context-dependent memory (memory for information is worse when tested in a new context than when tested in the original context in which it was learned; e.g., Kole, Healy, Fierman, & Bourne, 2010; Smith & Vela, 2001). An important corollary to this procedural reinstatement principle is that specificity (limited transfer) occurs for tasks based primarily on procedural information, or skill, whereas generality (robust transfer) occurs for tasks based primarily

on declarative information, or facts (Healy, 2007; Healy et al., in press). Thus, for skill learning, retention is strong but transfer is limited, whereas for fact learning, retention is poor but transfer is robust.

As mentioned above, an important distinction to keep in mind in any discussion of training is the difference between implicit and explicit learning. Implicit learning usually refers to the acquisition of skill or procedures, which is often accomplished by repetition and practice and does not necessarily involve intention. Furthermore, the skill that results from implicit learning is not necessarily conscious and can be applied automatically. In contrast, explicit learning usually refers to the acquisition of facts or new associations (also referred to as declarative knowledge). Explicit learning is generally accomplished intentionally by instruction, is applied consciously, and may not require repetition for its acquisition. This distinction between explicit and implicit learning provides an alternative formulation for the procedural reinstatement principle: Facts that are acquired explicitly may be rapidly forgotten; however, if they are available, they transfer broadly across new situations (e.g., Postman & Underwood, 1973). In contrast, skills that are acquired implicitly are well retained but transfer minimally to new situations (Ivancic & Hesketh, 2000; Lee & Vakoch, 1996; Maxwell, Masters, Kerr, & Weedon, 2001). It should be noted, however, that explicit learning might, with extended practice, become implicit, as in the proceduralization (or knowledge compilation) hypothesis of Anderson’s (1983) ACT-R theory.

Guideline: Trainers should reinstate the conditions of study as closely as possible when taking a test or performing in the field. If trainers are able to anticipate the test or field conditions, then they should modify their study conditions to match them. To make learning generalizable, training should be related to explicit declarative facts, whereas to make learning durable, training should be related to implicit procedural skills.

2. Specificity of training

Instructors often assume that teaching a primary task without extraneous secondary task requirements will benefit the learning process. However, if such secondary task requirements exist in the field, then use of this training method will not provide optimal transfer to field performance. Research has shown that to be effective, training must incorporate the complete set of field task requirements, including all secondary task requirements imposed in the field. This effect works both ways. That is, training with extraneous secondary task requirements will not be optimal if field performance does not include those requirements. In general, learning is highly specific to the conditions of training. This observation follows from both the specificity of training principle (retention and transfer are depressed when conditions of learning differ from those during subsequent testing; Healy & Bourne, 1995; Healy et al., 1993) and the functional task principle (secondary task requirements are often integrated with primary task requirements during learning, resulting in the acquisition of a single functional task rather than two separate tasks; Healy, Wohldmann, Parker, & Bourne, 2005; Hsiao & Reber, 2001). *Guideline*: For optimal performance, the entire configuration of task

requirements during training, including secondary as well as primary tasks, needs to match those in the field as closely as feasible.

C. Principles relating to task parameters

Training can vary along a number of dimensions depending, for example, on the task demands and properties. Certain training principles follow from variations in these task characteristics. The most well-established of these principles are described next, grouped by the task parameters entailed.

1. Spacing

When training new knowledge or skills involves repeated practice trials, learning is more efficient when rest intervals are interpolated between trials (i.e., spaced or distributed practice) than when the trials are administered without rest intervals (i.e., massed practice) (see, e.g., Bourne & Archer, 1956; Underwood & Ekstrand, 1967). A related spacing effect involves the separation of repetitions of a given item within a list of items (see, e.g., Glenberg, 1976; Hintzman, 1974). Although usually some rest between repetitions improves performance, the rest interval cannot be increased indefinitely. There is an optimal rest interval for at least some tasks (Bourne, Guy, Dodd, & Justesen, 1965), but more research needs to be done to determine the generality of this effect. With respect to retention of the learned material, this spacing effect does not always hold when the retention interval (interval between the last repetition and the test) is very short. Generally, the advantage of spacing holds for pure lists with a single interval as well as for mixed lists including intervals varying across different items (Kahana & Howard, 2005). All of this work is based on single-session training paradigms with short spacing and retention intervals.

In a different paradigm, Bahrick (1979) used long spacing intervals separating learning sessions and long retention intervals between the end of learning and final testing to study the acquisition of English-Spanish vocabulary pairs. Bahrick systematically varied the interval between practice sessions (intersession interval) during learning from 0 to 30 days, and he tested performance 30 days after the last learning session. He found that the level of performance on the final test session depended more on the spacing between learning sessions than it did on the level of performance achieved in the final learning session. Unlike findings from experiments with short intervals between practice trials or items (cited above), which generally show an advantage for spaced practice, performance on the final learning session of Bahrick’s study was greatest when the intersession intervals were shortest, but performance on the final test session was highest when the intersession intervals were longest (so that they resembled the retention interval). Bahrick, thus, concluded that for optimal knowledge maintenance, practice should be spaced at intervals approximating the length of the eventual retention interval. Bahrick and Phelps (1987) and Bahrick, Bahrick, Bahrick, and Bahrick (1993) confirmed this conclusion in studies involving retention intervals up to 50 years. For a summary of this work, see Bahrick (2005; but see Roediger, 2008).

More recently, Pashler, Rohrer, Cepeda, and Carpenter (2007) looked at the effects of varying the intersession interval (ISI). They showed strong effects of spacing over long retention intervals (RIs). In addition, test performance after a given RI was found to be optimal when the ISI was intermediate in value. Making spacing longer than optimal was, however, less harmful to retention than making it shorter than optimal. These authors suggest that it is more effective to use an ISI of several months or years than to use shorter intervals when retention is tested after a delay of several years. They found that the same spacing principles are applicable to some forms of mathematical skill learning, but not to perceptual categorization tasks. Kornell and Bjork (2008) showed that the induction of painter's styles was aided by spacing exemplars of each painter as compared to massing the exemplars. This result was surprising in that it had been thought that massed presentation would enable the subjects to more easily discover the similarities of the paintings by each painter. The authors proposed a new hypothesis that involved differentiating the individual styles of each painter, as opposed to highlighting the similarities of one painter's works. Seeing the different painters' paintings interleaved forced subjects to differentiate better among the various painters.

Arithmetic problems can often be solved either by calculation or by direct retrieval of the answer from memory. Calculation usually requires several steps and thus takes longer. Rickard, Lau, and Pashler (2008) found that with practice on the same problems direct retrieval from memory tends to replace calculation of the answer. They also discovered that in the training session this transfer from the slower calculation to the faster direct retrieval occurred sooner when the specific problems were spaced closer to each other (fewer other problems in between) than they did when they were spaced farther away (more other problems in between). However, in a test session days later the opposite result was found. These results are also consistent with the training difficulty principle, which states that a condition that causes difficulty during learning is beneficial to later retention and transfer (see below).

Rickard, Cai, Rieth, Jones, and Ard (2008) looked at the widely believed idea that sleep consolidation enhances skilled performance (see Marshall & Born, 2007; Stickgold, 2005; Walker, 2005; Walker & Stickgold, 2004, 2006). Rickard et al. used a sequential finger-tapping task and did find results that fit with sleep enhancement when data were averaged in the usual manner, that is, when 1 min or more of task performance at the end of the training session was compared with performance in the test session. This averaging could cause an illusory enhancement effect. However, they identified four aspects of the design and analysis not related to sleep consolidation that could lead to this enhancement effect. When they controlled for these factors in the data analyses or in the design, they did not find sleep enhancement as measured by either accuracy or reaction time. Rickard et al. concluded that sleep does not enhance learning for the explicit motor sequence task they used. They propose that the effects can be explained in terms of performance fatigue. With a long training session substantial fatigue builds up and creates an apparent asymptote in learning. This fatigue dissipates between sessions, which results in an apparent sleep enhancement effect on the test. This is the same effect that can be observed in spaced practice (as opposed to massed practice) in which the fatigue buildup dissipates during the space between practices. Rickard et al. suggest that

although sleep might not produce performance enhancement, it might provide a protection from forgetting (or a type of stabilization). This protection could be achieved in either an active or a passive manner. The active form would involve a mechanism that complements waking consolidation to produce stabilization. Thus, the mechanism involved in sleep consolidation might have a unique role distinct from that involved in waking consolidation. On the other hand, sleep might serve to protect against forgetting in a passive way. Thus, sleep might allow a more efficient operation of time-based consolidation because no new motor learning would occur during sleep that would interfere with any ongoing consolidation (see Wixted, 2004, for a similar explanation for sleep effects involving tasks using declarative memory).

Guideline: For optimal benefits from training, repeated practice on particular items or responses should be spaced in time. The amount of spacing (length of the time interval between repetitions) should be related to the amount of time that is likely to pass between training and eventual testing. Generally, it seems desirable to match the time between repetitions during training to the time between training and test.

2. Feedback

Two distinct questions have been asked about the effects of feedback: what form it should take and when to provide it.

a. What kind of feedback to provide

What type of feedback to provide is also a crucial issue for optimizing training and retention of knowledge and skills (Schmidt & Bjork, 1992). Trial-by-trial feedback has been shown to facilitate rate of learning in many tasks, possibly by motivating participants to set increasingly higher standards of performance or by identifying errors and how to correct them. But, if participants have a good sense anyway of how well they responded, then trial-by-trial feedback might be distracting, resulting in inferior performance on later acquisition trials, on retention tests, or on tests with tasks requiring slightly different responses. In such circumstances, periodic summary feedback, given only on some proportion of training trials, is often a more effective procedure for promoting long-term retention than is trial-by-trial feedback (see, e.g., Schmidt, Young, Swinnen, & Shapiro, 1989, for illustration of this finding in a ballistic timing task). Indeed there is some suggestion in the literature that the amount of feedback given during acquisition can be gradually reduced or faded without serious or adverse effects on acquisition performance and at the same time produce beneficial effects on long-term retention (Schmidt & Bjork, 1992). Other studies suggest, however, that any effects of feedback during training might not persist into later testing for retention (Bourne, Healy, Pauli, Parker, & Birbaumer, 2005).

b. When to provide feedback

In a declarative memory task, such as vocabulary learning, feedback is most effective for learning and retention when it serves to correct erroneous responses.

Pashler, Cepeda, Wixted, and Rohrer (2005) examined the effects of feedback to the learner in a foreign vocabulary-learning task. Different groups of subjects were provided with (a) simple right/wrong feedback after every learning trial, (b) feedback that signaled the correct responses, or (c) no feedback at all. They found that feedback had a facilitative effect on learning and on subsequent delayed recall of newly learned vocabulary but only when the feedback was provided after an incorrect response. Feedback had no benefit on correct response trials even when those responses were given with low confidence. On the other hand, in a concept-learning task Bourne, Dodd, Guy, and Justesen (1968) found facilitative effects of feedback on both correct response and incorrect response trials. The difference between the effects of feedback on the two types of tasks might relate to differing task requirements and the fact that there is an underlying abstraction in the concept-learning task used by Bourne et al. but not in the verbal associative task used by Pashler et al. Thus, in the concept-learning task, feedback serves to either confirm or disconfirm on every trial the learner's current hypothesis about the underlying concept, whereas in the verbal associative task, feedback on any given trial pertains only to a specific association, which has already been formed on the correct response trials. In a task different from both vocabulary and concept learning, namely recall of trivia, Smith and Kimball (2010) found facilitative effects of feedback following correct responses as well as errors, but these effects depended on the introduction of a delay before feedback is presented. Thus, the issue of task differences needs to be clarified in future research.

In a study of message comprehension in a navigation task, Schneider, Healy, Buck-Gengler, Barshi, and Bourne (2007) found that training with immediate feedback led to worse performance at test than did training with delayed feedback. These results suggest that immediate feedback, even when it provides supplemental information otherwise not available, might not always be desirable. In some cases, it might interfere with memory because of the interruption of the processing stream that supports learning. Further along those lines, Butler, Karpicke, and Roediger (2007) found not only that delayed feedback was better than immediate feedback for long-term retention but also that a longer delay (1 day) was better than a shorter delay (10 min.). An explanation for the benefit of delaying the presentation of feedback after a test is that feedback then serves as an additional spaced presentation of the information (see above). Immediate feedback is more consistent with massed presentations. Pashler et al. (2007) agree that immediate feedback may not be optimal and that delayed feedback may provide spaced practice especially after correct answers. Likewise, Wulf, Shea, and Whitacre (1998) point out that, in learning a motor skill, knowledge of results (KR) given too frequently or too quickly after the response might improve performance at the time of practice but impair later performance relative to learning a motor skill with KR that is given less frequently or after a delay (Gable, Shea, & Wright, 1991; Schmidt et al., 1989; for a review, see Schmidt, 1991).

Guideline: Informative feedback to the trainee is almost always desirable, especially early in the training process. However, the frequency of feedback can be reduced as the trainee acquires the required knowledge and skill. In fact, reduced feedback during training often facilitates long-term retention. Feedback with respect to

erroneous responses is generally more effective than feedback with respect to correct responses, and delayed feedback is sometimes preferable to immediate feedback, presumably because of a spacing effect (see above).

3. Rehearsal

a. Mental versus physical rehearsal

Often a skill-based task can be practiced either physically (i.e., by making the actual required responses) or mentally (i.e., by merely imagining the required responses). A number of studies have reported no benefits of mental practice (e.g., Shanks & Cameron, 2000), whereas other studies have reported benefits on tasks that are largely cognitive, but not on tasks that are largely motoric (e.g., Driskell, Copper, & Moran, 1994; Minas, 1978). But other studies have shown clear benefits to performance after mental practice even for motoric tasks (e.g., Kohl & Roenker, 1983), and Decety, Jeannerod, and Preblanc (1989) reported behavioral similarities between mental and physical practice of walking, either blindfolded or by imagination, to specified locations at varying distances. Furthermore, Wohldmann, Healy, and Bourne (2007) demonstrated in the context of a simple perceptual-motor laboratory task that some aspects of mental and physical practice are similar behaviorally in that mental practice is just as effective as physical practice both for learning a new motor skill and for maintaining a previously learned motor skill across a 3-month delay. In fact, Wohldmann, Healy, and Bourne (2008a) established that mental rehearsal is in some circumstances better than physical rehearsal in promoting the acquisition, durability, and transferability of perceptual-motor skill because mental rehearsal does not suffer from interference effects attributable to physical movements.

b. Fixed versus expanding rehearsal

The studies of spacing effects reviewed above all used fixed intertrial intervals during training. Landauer and Bjork (1978) suggested that constant intervals, regardless of their length, might not be optimal for learning and retention. They examined a training procedure in which the intervals between test trials gradually increased during learning. This expanding rehearsal procedure produced greater eventual performance than did a rehearsal procedure with uniform intervals between tests. The positive effects of expanding rehearsal have been replicated by Cull, Shaughnessy and Zechmeister (1996; see also Morris & Fritz, 2000), but there have been some failures to replicate (Cull, 2000). In fact, Karpicke & Roediger (2010) suggested that the positive effects of expanding rehearsal might be due to the greater amount of spacing under expanded, as opposed to fixed, rehearsal conditions. When the amount of spacing was controlled, the difference between fixed and expanding conditions disappeared in their study. However, a recent study by Storm, Bjork, and Storm (2010) found conditions under which expanding rehearsal is effective, namely those involving material that is highly vulnerable to forgetting. In any event, an interesting possible extension for future experimental study is to expand the intervals between training sessions following the work of Bahrick (1979, 2005) summarized above. Although Bahrick found it optimal to

match the interval between training sessions to the retention interval separating the last training session and the test session, it may be instead that optimal performance occurs with an expanding set of intervals between training sessions, with only the last equal to the retention interval.

Guideline: Type and scheduling of rehearsal opportunities can have important impacts on the acquisition, retention, and transfer of knowledge and skill. In general, mental rehearsal should be employed whenever physical practice is difficult or impractical. Also, expanding rehearsal might be considered as a possible strategy, if there is sufficient time during training to allow for the spacing that is entailed, but the supporting empirical evidence is still lacking.

4. Testing

Tests are usually thought of as performance assessment tools, but there is increasing evidence that people learn from taking tests often as much or more than they learn from pure study. This phenomenon has been referred to as a “testing effect” (Carpenter & DeLosh, 2005; Izawa, 1992; McDaniel & Fisher, 1991). Specifically, the testing effect is the advantage in retention for material that is tested relative to material that is presented for additional study. A number of theoretical explanations have been proposed for the testing effect (see Dempster, 1996, and Roediger, 2009, for reviews), such as those involving the amount of processing and retrieval practice. This effect has been demonstrated for both semantic (e.g., words) and nonsemantic (e.g., unfamiliar faces) materials (Carpenter & DeLosh, 2006) (but see Roediger, 2008).

Marsh, Roediger, Bjork, and Bjork (2007) found that it is detrimental to students to be exposed to plausible wrong answers on a multiple-choice test, even if the students choose the right answer. In addition, multiple-choice lures may become integrated into the learners’ more general knowledge and lead to erroneous reasoning about concepts. However, the authors believe that the overall positive effect of testing outweighs any negative consequences, and they show in several studies that the learning of lure answers was balanced by a decrease in other wrong answers on the final tests. Marsh et al. make three suggestions to help prevent the problem of lures being produced on a later test. The first suggestion is to give immediate feedback. Immediate feedback should reduce the chance of producing on a subsequent test a previous multiple-choice lure (Butler & Roediger, 2006) (but see the discussion above concerning immediate vs. delayed feedback). The second suggestion follows the SAT II’s practice of providing a “don’t know” option and giving a penalty for any wrong answer. Being given the option of “don’t know” and being penalized for wrong answers should significantly reduce lure production on a subsequent test involving cued recall. The third suggestion is to alter across exams how concepts are tested. A change from a multiple-choice question requiring a definition to a cued-recall question requiring application should serve to reduce, although perhaps not eliminate, the negative consequences of multiple-choice lures.

Pashler et al. (2007) point out that the testing effect has been found for various types of tests and materials. Specifically, the effect is evident for free recall (e. g., Allen, Mahler, & Estes, 1969; Carpenter & DeLosh, 2006) and cued recall (Carrier & Pashler, 1992) and for face-name associations (Carpenter & DeLosh, 2005), definitions (Cull, 2000), and general knowledge facts (McDaniel & Fisher, 1991). They also found that covert retrieval practice, a form of mental rehearsal, in which subjects are asked to retrieve without providing an observable response, enhances learning. McDaniel, Roediger, and McDermott (2007) illustrated the testing effect in real life, that is, in an actual course at a university. They found that providing short-answer and multiple-choice tests initially, compared to providing no tests initially, significantly aided performance on a subsequent test. They also found that short-answer tests (requiring production or recall) were more helpful to later test performance than were multiple-choice tests (requiring only recognition), even when the later tests invovled multiple-choice questions. Finally, they found that short-answer tests were more effective than focused study, especially when those tests involved corrective feedback.

Note that the testing effect has been examined primarily in declarative leaning tasks, where it is possible to separate pure study from test performance. In skill learning tasks, study and tests are usually integrated into the trial-by-trial acquisition procedure, with each trial necessarily including a testing component. The testing effect is really, thus, not directly applicable to skill learning although mental practice (or even observation) might be considered an analogue of studying without testing.

Guideline: A lot of learning occurs during test taking. Therefore tests should be embedded in the training process whenever possible.

5. Overlearning

Training usually ends when the trainee reaches some predesignated performance criterion, such as one or more error-free training trials. Overlearning refers to practice beyond the performance criterion (Pashler et al., 2007). It has been found that overlearning, relative to less practice, improves later performance (Krueger, 1929). Consequently, overlearning has been proposed as a useful, general strategy when long-term retention is the goal (Driskell, Willis, & Copper, 1992). However, overlearning might not be an efficient way to strengthen acquired knowledge and skill. For example, in a study by Rohrer, Taylor, Pashler, Wixted, and Cepeda (2005) subjects were taught novel vocabulary pairs. They saw each word pair either 5 or 10 times. After 1 week, the subjects who saw the pairs 10 times showed a substantial benefit over the subjects who saw the pairs 5 times, but the difference had disappeared after 4 weeks. Rohrer and Taylor (2006) conducted a similar study using a new math skill. One group of subjects had three times the number of practice problems but no difference was found after either the 1-week or the 4-week retention interval. Thus, Pashler et al. conclude that for long-term memory, overlearning seems to be inefficient as a training technique. They point out, however, that in some cases overlearning might be the only alternative when a skill needs to be performed with absolutely no errors at a much later time (e.g., performing CPR or landing a space shuttle). They also say that, even when retrieval accuracy is at

ceiling, overlearning might improve speed of responding (e.g., Logan & Klapp, 1991), and speedup could be useful when rapid responding is a prime consideration.

A related phenomenon has been identified as “the failure of further learning effect.” This effect was first demonstrated by Kay (1955) and Howe (1970), and subsequently studied by Fritz, Morris, Bjork, Gelman, and Wickens (2000). Repeated studying of text passages presented out loud to subjects yields little new learning beyond that attained in the initial study period, even though there is much additional information to be learned and the learning is spaced rather than massed. An explanation offered by Fritz et al. for this effect is that the learner develops a schema (or mental summary) reflecting his or her comprehension of the text as a result of the first study episode and that schema creates some resistance to improving learning after it has been established. They also interpret the findings in terms of the distinction between “given” (i.e., known) and “new” (i.e., yet to-be-learned) information (Haviland & Clark, 1974), with the hypothesis that learners neglect information that they consider to be given (because it was included previously) even though they have not been able to recall it.

Guideline: Overlearning is recommended as a training technique only when training time is not severely limited and when it crucial to have the strongest possible representations of knowledge and skill.

6. Task difficulty

Interference is a source of difficulty in training that occurs when conditions allow incorrect answers to come to the trainee’s mind, along with the correct answer, thereby requiring the trainee to choose the correct answer from among several alternatives. Increasing interference during training has been shown to impede training speed but ultimately to enhance the durability and flexibility of what is learned. For example, mixing material across categories during training, as opposed to grouping the material by category, enhances interference, which may inhibit initial acquisition, but should yield better retention and transfer. In fact, it has been shown that many things that make learning difficult (not just interference) facilitate transfer to a new task as well as long-term retention of the original task. This recommendation follows from both the effects of contextual interference (interference during learning facilitates later retention and transfer; Battig, 1972, 1979; Carlson & Yaure, 1990; Lee & Magill, 1983; Schneider, Healy, & Bourne, 1998; Schneider, Healy, Ericsson, & Bourne, 1995; Shea & Morgan, 1979; but see Wulf & Shea, 2002, for some exceptions) and, more generally, the training difficulty principle (generally, any condition that causes difficulty during learning facilitates later retention and transfer; Schmidt & Bjork, 1992; Schneider, Healy, & Bourne, 2002; but see McDaniel & Einstein, 2005, and Young, Healy, Gonzalez, Dutt, & Bourne, in press, for some qualifications).

Not all sources of difficulties during training are desirable, however (see Bjork, 1994). McDaniel and his colleagues (McDaniel & Butler, in press; McDaniel & Einstein, 2005) argue that difficulties introduced during training are facilitative only when they

cause the learner to engage in task-relevant processes that otherwise would not take place.

Guideline: Counter to intuition, trainers should consider introducing sources of interference into any training material. If durable retention and flexible transfer are the goals of training, then mixing materials during training is advisable for most learners. Trainers might consider enhancing the difficulty of training exercises in other ways as well with the caveat that task-relevant cognitive processes must be engaged.

7. Stimulus-response compatibility

Cognitive skills can be divided into three stages: (a) perception of the stimulus, (b) decision making and response selection, and (c) response execution (Proctor & Dutta, 1995). The most ubiquitous phenomenon observed in the second stage of skill acquisition is the effect of stimulus-response compatibility (Fitts & Deininger, 1954; Fitts & Seeger, 1953; Proctor & Vu, 2006). This effect reflects a difference in performance attributable to the mapping of individual stimuli to responses, such that performance is best when the stimulus set and the response set are configured in a similar way and each stimulus is mapped to its corresponding response (e.g., left-right stimulus locations are mapped to left-right responses). Stimulus-response compatibility effects have been extensively studied using stimuli and responses with spatial properties, but they occur for any dimension of similarity between stimuli and responses. The detrimental effects of incompatibility are not easily overcome, even after extensive practice (e.g., Dutta & Proctor, 1992). *Guideline*: It is important to maintain stimulus-response compatibility during training to avoid the prolonged, detrimental effects that incompatibility can have on performance.

8. Seeding

When tasks require having a certain type of quantitative knowledge, providing a small number of examples, called *seeds*, is often sufficient knowledge to encompass an entire domain. For example, for a quantitative estimation task (e.g., estimating the distances between geographical locations), providing a small number of specific relevant quantitative facts can greatly improve overall estimation ability. A small number of sample distances is extremely beneficial not only to immediate estimation but to estimation performance after long delays. This recommendation follows from the seeding effect (Brown & Siegler, 1996, 2001; Kellogg, Friedman, Johnson, & Rickard, 2005; LaVoie, Bourne, & Healy, 2002).

However, seeding might not work in all cases. For example, in a study simulating scanning by airport screeners (TSA agents) (Smith, Redford, Washburn, & Taglialatela, 2005), when the same targets were repeated, the subjects could recognize familiar targets but had great difficulty generalizing to new or unfamiliar targets. Specifically, performance improved as test images repeated but dropped sharply when unfamiliar targets from the same categories were added. Thus, subjects relied on familiarity and had difficulty using category-general information. These results suggest that seeding effects

might be limited to certain domains such as those involving quantitative estimates.

Guideline: Seeding (training on a few specific examples of a selected domain) can be effective but should be used judiciously in non-quantitative domains, based on the likelihood of seeding effects in those domains.

9. Serial Position

Better memory has been found for the initial and final items in a to-be-learned list of items (Nipher, 1878). This bow-shaped serial position function, with both primacy and recency components, is found at the start of learning but diminishes as repeated trials on the same material are given (Bonk & Healy, 2010). The same effect is observed for short lists (as few as 4 items) and long lists (40 items or more), for tasks that require item learning or response-sequence learning, and for both immediate recall and serial learning. The relative magnitude of primacy and recency effects differs depending on many variables, especially the testing procedure. In any event, the items in the middle of a list are at a disadvantage when it comes to both short-term memory and long-term acquisition. Thus, training will require more practice on items in the middle of a list than on those at either end. *Guideline*: For tasks that require training on a sequence of informational items or responses, the trainer should place greater emphasis on items in the middle of the sequence than on those at the beginning or end.

D. Principles relating to individual differences

Training principles are likely to apply unequally across individuals and to the same individual in different circumstances. There are some systematic inter- (between) and intra- (within) individual differences that should be considered in the design of training routines.

1. Zone of learnability

As an example of an important individual difference that applies both among different individuals and within the same individual at different times is the “zone-of-learnability.” The zone-of-learnability refers to material that contains information that is a little beyond what a particular student already knows, neither too close to nor too far away from what is already known (Wolfe, Schreiner, Rehder, Laham, Foltz, Kintsch, & Landauer, 1998). People learn most efficiently when the material to be learned is within their zone of learnability. This principle has also been referred to as the “Goldilocks hypothesis” (implying that the material to learn is just right, neither too simple nor too difficult). Related to this principle is the established finding that learning from text is better if the learner has appropriate background knowledge (e.g., Means & Voss, 1985; Moravcsik & Kintsch, 1993), so that a central feature of learning from text is linking up the information in the text to the reader’s pre-existing knowledge. That is, new information in a text needs to be integrated with the reader’s pre-existing knowledge. If there is no relevant information base, then the integration cannot take place, and no learning will occur. For optimal learning, text difficulty should be matched to the

student's level of background knowledge, so that easier texts should be used for students with a lower level of prior knowledge. According to the zone-of-learnability principle, optimal learning occurs when the text provides some, but not too much, new information.

One way to establish the zone of learnability in a group of students is to use the newly developed clicker technology, which is based on periodic multiple-choice testing within an ongoing lecture. The technique makes use of a personal response system provided to each student with which the student responds to the multiple-choice probe questions. When most students respond correctly, the trainer can assume that the material presented is well within the students' zone of learnability and can move forward. If most students respond incorrectly, the trainer has reason to assume the material is not yet within the zone of learnability so that clarification or repetition is necessary. Evidence to date on the clicker technology is limited but promising (Anderson, Healy, Kole, & Bourne, 2010; Mayer et al., 2008).

When training involves learning information from text (e.g., from written instructions), it is also important to consider the type of text to be used. In general, coherent text (which is harmonious and logically consistent) is advisable. However, the readers' existing domain knowledge determines whether they will benefit from a coherent text (McNamara & Kintsch, 1996; McNamara, Kintsch, Songer, & Kintsch, 1996). Readers with low knowledge learned more effectively with high-coherence text, whereas, counter to intuition, readers with high knowledge benefited from a low-coherence text according to some measures. Specifically, text coherence had little effect for high-knowledge readers' memory in terms of their recall and accuracy on comprehension questions that were derived from a single idea in a text (rather than those derived from a relation between several ideas expressed in the text). But there was a clear benefit to high-knowledge readers for low-coherence text in terms of measures reflecting the readers' understanding of the concepts conveyed in the text. In summary, only low-knowledge readers show a benefit from reading a high-coherence text. High-knowledge readers actually show more understanding of the relevant concepts after reading a low-coherence text (McNamara, 2001), which is consistent with the concept of zone-of-learnability.

Guideline: It is important for the trainer to be sensitive to the trainee's current level of knowledge in the relevant domain and to attempt to find learning materials that are appropriate to that level of knowledge. To establish the level of knowledge of a group of trainees, the newly developed clicker technology should be considered.

2. Strategy variation

Trainers need to be sensitive to the fact that different strategies might be optimal for different learners, at different stages of skill or knowledge acquisition, and with different learning material. For example, some materials might be best mastered by rote learning or memorizing specific instances, whereas other materials might benefit from a more abstract rule-learning approach. Instance-based strategies are preferred and lead to more efficient performance in simple tasks, whereas rule-based strategies are optimal in more

complex tasks (Bourne et al., 1999; Bourne, Healy, Kole, & Raymond, 2004). Rules might be particularly important to formulate and use when the number of instances to be dealt with challenges or exceeds available memory and when the individuals lack confidence in their ability to remember instances (Touron, Hoyer, & Cerella, 2004). Further, rules tend to be more durably represented in memory than are instances. When performance after a delay is of crucial concern, then training procedures need to emphasize rule-based strategies, rather than instance-based strategies, because the rule will be better retained than instances across a delay (Bourne, Healy, Kole, & Graham, 2006; Bourne, Parker, Healy, & Graham, 2000). Although these effects hold in the aggregate, individuals vary in the extent to which they rely on instance memory versus a rule-based strategy, some individuals persisting in a rule strategy long after others have switched to memory-based responses (Bourne, Raymond, & Healy, 2010; Rickard, 2004). *Guideline*: When the most effective strategies for a given task are known, instructors would be advised to adopt procedures that can bring these strategies forward earlier than usual in the training process.

3. Chunking

When a series of items (e.g., a list of words) is presented, subjects can usually recall about seven of them, which is called the immediate *memory span*. Classic research has shown that it does not matter much what the items are; they can be digits, letters, words, or even phrases. The limit is always about seven. This finding gives rise to the idea that people can combine presented material into units of different sizes, which are called "chunks" (Miller, 1956), and that they can recall about seven chunks, regardless of what is in them. This result suggests that a good memory strategy is to try to find ways to chunk material that needs to be remembered. Indeed it is possible, with deliberate practice that builds on existing chunks of digits such as dates and running times, to increase the digit span to a very large number (Ericsson, Chase, & Faloon, 1980). This expansion of memory is not without limits. As the size of the unit to be remembered increases, the number of chunks that can be recalled shrinks. Some people have suggested that, at least with very large chunks, the immediate memory span is closer to three (Broadbent, 1975; Cowan, 2001, 2010). For example, in experiments simulating communication between pilots and air traffic controllers as to navigation in space, Barshi and Healy (1998, 2002) found that subjects could recall up to three commands with very little error. Beyond that number, however, recall performance fell off dramatically, although practice was able to offset the decline to some extent. *Guideline*: Trainers should encourage a chunking strategy wherever possible for acquiring and recalling large amounts of material. Furthermore, when providing a sequence of information to be recalled, trainers should divide the material into segments that include no more than three units or steps at a time.

IV. Partially established training principles

Some training principles are not fully established at the present time and require additional supportive research. Important partially established training principles will now be reviewed, under the same four categories as used above for the well established

principles: (a) resource and effort allocation, (b) context effects, (c) task parameters, and (d) individual differences.

A. Resource and effort allocation

1. Focus of attention

It is possible for a learner to deploy or focus attention in various ways during training. Furthermore, a learner might be instructed effectively about how to focus attention. Some studies have compared an external focus of attention (i.e., attention to the results of a movement) of learned motor skills to an internal focus of attention (i.e., attention to the body movements themselves). That research has consistently found, at least after some initial training, that there is an advantage for the external focus of attention with respect to learning, retention, and transfer of motor skills (McNevin, Shea, & Wulf, 2003; Shea & Wulf, 1999; Wulf, McNevin, & Shea, 2001). This result is explained by the constrained action hypothesis, according to which well developed motor skills are represented by automatic mechanisms within the body that are impaired by conscious attention to them (Beilock, Bertenthal, McCoy, & Carr, 2004). *Guideline*: Trainers should encourage learners to adopt an external focus of attention on the target of their movements rather than on the bodily movements themselves.

2. Strategic use of knowledge

When trainees need to learn a large amount of new information, that information should be related to their existing knowledge. Previously acquired knowledge can be used as a structure for organizing otherwise unrelated facts even when the facts themselves fall outside the domain of existing knowledge. For example, if trainees know a lot about baseball, they can use that knowledge to organize and, thus, quickly learn a large set of facts about members of their crew. The idea is to associate each member of the crew with a famous individual from the baseball domain. Although additional associations might seem to complicate the task at hand, connections to existing knowledge will enhance performance both in terms of accuracy and speed of responding with the new information, following the strategic-use-of-knowledge principle (learning and memory are facilitated whenever pre-existing knowledge can be employed as a mediator in the process of acquisition; Healy, Shea, Kole, & Cunningham, 2008; Kole & Healy, 2007; Van Overschelde & Healy, 2001). Chunking is a special case of the strategic use of existing knowledge (see above). *Guideline*: Trainees should be instructed to use their previously acquired knowledge when learning a new set of facts, even if the existing knowledge seems irrelevant to the new facts.

3. Cognitive antidote to fatigue and boredom

Prolonged work on a given task often results in deterioration of performance, despite ongoing skill acquisition. It has been found that prolonged work sometimes produces an increasing speed-accuracy tradeoff in performance, such that accuracy declines over trials while at the same time response speed improves (Healy et al., 2004;

see the discussion of speed-accuracy tradeoffs above). The deterioration is attributable to fatigue, task disengagement, or boredom on the part of subjects. This deterioration can be counteracted by the introduction of a simple cognitive requirement on each response. For example, subjects might be required to make a simple computation before each response or to alternate terminating keystrokes after each response (Kole et al., 2008). Under these conditions, the speed-accuracy tradeoff is eliminated; that is, the decline in accuracy disappears although responses continue to speed up across practice trials. These results have led to a cognitive antidote training principle (the introduction of cognitive activities can counteract fatigue, task disengagement, and boredom effects, resulting in performance maintenance or even improvement during sessions of prolonged work). *Guideline*: Instructors should consider adding a cognitive component to a routine task on a trial-by-trial basis to avoid disengagement and boredom. This added cognitive component is likely to be most effective when it is relevant to the ongoing training task or simple in nature.

B. Context effects

1. Part-task training

Under certain conditions part training (training only a part of a task before training the whole task) is more effective than whole training (training the whole task from the beginning). Part training can either involve forward chaining (when the initial segment of a task is trained first) or backward chaining (when the final segment of the task is trained first). For complex tasks that can be divided into components, the conditions for part-training superiority appear to be a function of the organization of subtasks. Complex tasks can be organized in at least two different ways: A segmented task contains parts that are performed sequentially, whereas a fractionated task contains parts that are performed simultaneously. Part-task training is most beneficial when performing a backward-chaining procedure in a segmented task (but see Peck & Detweiler, 2000, for a demonstration of the effectiveness of a forward-chaining technique). Wightman and Lintern (1985) argue that the backward-chaining method is superior because there is a strong association between performance level on the terminal task and knowledge of results (i.e., the feedback resulting from task completion). The results of Marmie and Healy (1995) with part training using backward-chaining on a segmented task add support to this argument. In contrast, for a fractionated task, Adams and Hufford (1962) found that training first on only one procedure initially disrupted performance on the whole procedure. Marmie and Healy (1995) offer the following explanation: In both types of tasks, during the initial part-training phase, the trainee constructs independent procedural representations for each part of the whole task. When transfer to the whole task occurs, there is only a single interruption between the two parts in a segmented task but multiple interruptions in a fractionated task. Thus, the procedural representations can remain intact and independent only in a segmented task; in a fractionated task a new procedural representation must be established, which requires integration of the two parts, because the parts in that case are performed as an interlocking unit. In addition, findings described below suggest that segment difficulty as well as segment position in the sequence must be considered when designing a part-task training method.

Naylor and Briggs (1963) found support for the hypothesis that the relative efficiency of part-task and whole-task training is related to an interaction between task complexity and task organization. For an unorganized, complex task, they found that part practice surpassed whole practice in efficiency, but on all other combinations of task complexity and task organization, groups trained by the whole method were superior to progressive-part groups during transfer. Brydges, Carnahan, Backstein, and Dubrowski (2007) supported the view that a motor skill involving high organization and high complexity needs to be practiced under whole practice conditions, probably because moving from one skill to another in part practice changes the kinematic characteristics of each component. On the other hand, Anderson (1968) found that for first graders trained to solve concept-attainment problems, a whole-task group did not perform as well as a part-task group either on problems occurring at the end of training or on related problems presented subsequently in a retention test, but the two groups were equivalent on more dissimilar transfer problems. Newell, Carlton, Fisher, and Rutter (1989) suggest that the benefits of part-task training depend on the nature of the part task trained in prior practice. Only when the part-task training involves smaller subtasks with natural interconnected units will part-task training enhance whole-task skill acquisition. In agreement with this idea is Holding’s (1965) suggestion that practice subtasks should represent “small wholes” rather than isolated parts.

Guideline: Whether or not initial training of a complex task should involve only parts of that task depends on a number of task characteristics. Trainers need to be sensitive to these characteristics before deciding to use part-task training. Among the important factors are (a) forward versus backward chaining of the parts, (b) segmented versus fractioned nature of the whole task, and (c) dependency among the task components.

2. Easy-difficult ordering

Tasks can be divided into parts based on aspects of the stimuli involved, such as their difficulty. This division raises the question in part-task training as to which parts of a stimulus set should be trained first. When a task involving a stimulus set is trained incrementally, the question arises as to whether the easier or the more difficult stimuli in the set should be trained first. Pellegrino, Doane, Fischer, and Alderton (1991) found that initial training on a difficult subset of stimuli was beneficial relative to initial training on an easy subset of the stimuli in a visual discrimination task. (Related results in the training of motor skills have been reviewed by Schmidt and Lee, 1999.) According to Pellegrino et al. (1991; see also Doane, Alderton, Sohn, & Pellegrino, 1996; Doane, Sohn, & Schreiber, 1999), incremental training should begin with the part of the stimulus set that yields the most effective strategic skills. However it is not always the more difficult part that yields the optimal strategic skills. For example, Clawson et al. (2001) found that initial training on easy stimuli in a Morse Code reception task led participants to adopt an effective unitization strategy for representing codes, whereas initial training on difficult stimuli led to a less effective strategy in which individual elements were separately represented and then integrated.

Spiering and Ashby (2008), on a difficult perceptual categorization task, found that the effect of different training orders depended on the type of categories used. In rule-based category learning, processing through explicit reasoning is used. In this type of learning the rule is often easy to describe (Ashby, Alfonso-Reese, Turken, & Waldron, 1998). For category learning involving information integration, information from multiple stimulus components must be integrated before a decision is made. In that case, the optimal strategy is hard to describe (Ashby et al., 1998). When explicit reasoning can be used to learn the categories (rule-based task), the order in which training is presented does not matter. However, when the rule for categorization is hard to describe (information-integration task), difficult training first is the most effective method for learning.

A related issue that has been explored by Maxwell et al. (2001) is what they call *errorless learning* (see also Terrace, 1963, for earlier work with animals). For a motor skill, subjects should begin with the easiest task, where few if any errors are made, and progress to increasingly harder tasks to minimize the overall number of errors made. In golf putting, for example, learners should begin with a short-distance putt and progress to longer and longer putts. Maxwell et al. equate errorless learning with implicit learning and error-prone learning with explicit learning. It has been shown that skills that have been learned in an error-prone manner require more explicit, attentional resources than do skills learned in an errorless manner. Because there is less attention needed to perform the skill learned in errorless training, which seems to be more like implicit learning, distractions, such as a secondary task, cause less disruption. Hardy, Mullen, and Jones (1996) and Masters (1992) also found that skills learned implicitly are more immune to the negative effects of psychological stress (see the discussion above concerning the distinction between implicit and explicit learning).

Kern, Green, Mintz and Liberman (2003) found that errorless learning can be used to compensate for neurocognitive deficits relating to new skill acquisition and to rehabilitate persons with schizophrenia so that they can work effectively. In contrast, in other clinical research, in this case involving patients with phonological disorders, Gierut (2001) reported that training on the more difficult aspects of the phonological system yielded the greatest amount of generalization. This effect has also been shown with aphasic patients (e.g., Kiran & Thompson, 2003; Thompson, Shapiro, Tait, Jacobs, & Schneider, 1996) and in normal language development (e.g., Au, 1990; Eckman, 1977). These results indicate that there are limits on the benefits of errorless learning, at least in some domains, so that additional research is required to determine what order of components to use in training of a specific task.

Guideline: Whether or not training should begin with the easiest or most difficult components of a fractionated task depends once again on a number of task characteristics. Trainers need to be sensitive to these characteristics before deciding on the order of the subtasks. Among the important factors are (a) the parts that yield the best strategic skills, (b) explicit or implicit category definition in categorization task, (c) explicit or implicit learning in motor skills, and (d) the domain of knowledge and skill to be trained.

C. Task parameters

1. Variability of practice

Variable practice conditions (in which individuals train on a number of different tasks) typically yield better performance at transfer testing than do constant practice conditions (in which individuals train on a single task), even when testing is conducted on the same task as trained under constant practice. The benefits of variable practice were first recognized by Schmidt (1975) for discrete motor tasks and explained by him in terms of a schema theory, according to which variability promotes effective and general use of rules (schemata) relating external task requirements to internal movement commands. Wulf and Schmidt (1997) extended these findings to a continuous, feedback-regulated tracking task, and Schmidt and Bjork (1992) extended them further to tasks that do not involve motor learning, such as concept formation and text processing. Recently, Goode, Geraci, and Roediger (2008) also found that variable practice yielded superior transfer over repeated practice on anagram solutions. Specifically subjects practiced solving anagrams in one of three ways: (a) They repeatedly solved the exact anagram to be tested subsequently. (b) They repeatedly solved an anagram different from the one tested subsequently. (c) They solved different versions of the anagram tested subsequently. The third group, which used variable practice involving different anagram variations, performed better at test relative to the other two groups, even the group that practiced the exact same anagram included on the test.

Contrary to these findings, in a feedback-regulated non-tracking perceptual-motor task, Healy, Wohldmann, Sutton, and Bourne (2006) found that performance was worse for variable practice conditions relative to constant practice conditions involving the same task used during transfer testing. However, in a subsequent study involving the same perceptual-motor task, Wohldmann, Healy, and Bourne (2008b) found benefits of variable practice when subjects were given multiple targets under the same perceptual-motor reversal conditions, as opposed to being given the same targets in multiple perceptual-motor reversal conditions (Healy et al., 2006). Wohldmann et al. explained their findings by pointing out that if each reversal condition is assumed to involve a distinct configuration of responses (i.e., a distinct generalized motor program), practicing with multiple reversal conditions might not strengthen any one configuration, but practicing with multiple target locations within a single reversal condition should strengthen that configuration. In any event, an examination is warranted of the generality and boundary conditions of the variability of practice principle across task environments.

Guideline: Trainers should vary the conditions of practice to facilitate generalization of the trained skill. There are some limits, however, which involve how variability is introduced into the task. Current evidence suggests that variability is most effective when a single motor program is being learned so that variability applies to the context rather than the core program itself.

2. Modality effects

Presenting verbal information in the auditory modality generally aids memory for that information relative to presenting it in the visual modality (i.e., memory for verbal information is improved when it is heard rather than seen) (see, e.g., Gardiner, Gardiner, & Gregg, 1983). Explanations for this modality effect have included both the proposal by Penney (1989) that auditory and visual items are represented using different processing streams and the proposal by Mayer (2001) that there are two parallel channels for multimedia learning, the first including material that is visual/pictorial and the second including material that is auditory/verbal. By Penney's account, auditory presentation is superior because auditory material is automatically represented in an acoustic code that has a relatively long durability and large capacity, and that code is not available for visual material. Both auditory material and visual material are represented in a phonological code. In addition, visual material is represented in a visual code that has short durability and small capacity. By Mayer's account, spoken words are processed directly in the auditory/verbal channel, but written words are not processed directly in either channel even though written words are processed indirectly in both channels. Future research is needed both to verify that the auditory modality is superior in other domains (see Schneider, Healy, & Barshi, 2004, for one such recent verification in the domain of message comprehension), to clarify which of the alternative explanations is most consistent with the observed results, and to determine whether the same modality effects that apply to acquiring information also apply to the long-term retention and transfer of that information. *Guideline*: When the information to be learned is verbal (i.e., textual), then trainers should use auditory presentation rather than visual presentation to facilitate acquisition.

D. Individual differences

There are individual differences in abilities, performance, and preferences on any task. In fact, selection of trainees in the military and in industrial settings is generally based on tests of individual differences. The existence of individual differences suggest the possibility that people differ in their style or approach to performing particular tasks. Moreover, individual differences might change as a function of training. Both of these possibilities are considered in this section.

1. Learning styles

The idea that individuals differ in learning style is intuitive and popular (for a review see Kozhevnikov, 2007), but the evidence supporting these differences is weak. Pashler, McDaniel, Rohrer, and Bjork (2009) reviewed the evidence and concluded that it was not substantial enough to warrant any accommodations to training based on learning style. For example, studies comparing "visualizers" (individuals who prefer to work with pictorial materials) and "verbalizers" (individuals who prefer text-based materials) did not show convincingly that matching materials to purported learning styles resulted in any significant benefit, or in any aptitude-treatment interaction (ATI) (Massa & Mayer,

2006). *Guideline*: Until additional evidence is available, trainers should not attempt to tailor training to trainee preferences or alleged styles.

2. Effects of practice on individual differences

In addition to the amount of practice on a skill, individual abilities play a big part in the level of performance trainees achieve. Whether or not practice in a skill makes individuals more similar or more different depends on the task (Ackerman, 2007). For tasks that can be performed by most people, such as driving a car, consistent practice reduces the differences among people. Novices may start off with big individual differences in performance ability but have much smaller individual differences with practice. On more complex tasks, especially those that allow for successful performance by the use of differentially effective strategies that are beyond the capabilities of many, some people become very fast and accurate, whereas others remain at the novice level, leading to enhanced individual differences. Thus, for these complex tasks, the individual differences become larger with practice. After some level of automaticity is reached, two abilities are good predictors of performance following extensive practice: perceptual speed and psychomotor function.

For tasks that require declarative knowledge, performance levels depend on whether the tasks are “open” or “closed.” Closed tasks are limited to a finite domain of knowledge, whereas open tasks increase with complexity. For open tasks (but not for closed tasks) there is an increasing difference between the levels of the highest- and lowest-performing people. For tasks building on existing knowledge, individual differences in the extent of that knowledge are more important for acquiring new information than are individual differences in the capacity of working memory (Baddeley, 2007), or memory for recently presented material and actions (e.g., see Beier & Ackerman, 2005). It is also more important for learners to have a high level of knowledge in the relevant domain along with a high level of general, crystallized intelligence than to have a high level of fluid intelligence (reasoning ability) and working-memory capacity. Thus, the knowledge that an individual brings to the task is more important for determining what additional knowledge that individual can acquire later than is the individual’s working memory capacity, especially in areas such as health literacy or financial planning, but less so in areas such as math and physical sciences (see the discussion above on the strategic use of existing knowledge in learning new facts).

Guideline: Trainers should keep in mind that individual differences in performance might increase or decrease with practice depending on the complexity of the task to be learned and the relevant domain of knowledge. This fact suggests that the amount of training required to reach a criterion will differ across individuals, especially in complex tasks and in open tasks building on declarative knowledge.

V. Other considerations

There are other, miscellaneous factors, beyond those reviewed above, that need to be considered when developing a training program although they do not directly suggest specific training principles.

A. Global versus local processing

Under normal conditions the processing of global features dominates, or has precedence over, the processing of local features (Navon, 1977, 1991). In experiments involving large letters made up of small letters, individuals were usually faster to identify the large letter (global feature) than to identify the small letter (local features). An asymmetrical interference was also found in which there is interference in processing local features by global features but not the other way around (see, e.g., Kimchi, 1992; Kinchla, 1974). This asymmetrical effect has been shown to be sensitive to manipulations of various perceptual factors (see, e.g., Martin, 1979; Navon & Norman, 1983). The asymmetrical nature of global and local processing also depends on attentional factors, including, for example, whether attention needs to be divided between global and local targets (see, e.g., Robertson, Egly, Lamb, & Kerth, 1993; Ward, 1982). In fact, research has shown that global information affects the processing of local information even when the global information occurs in a stimulus that is unattended (e.g., Paquet, 1992). There is some evidence, however, that global information can be inhibited in cases requiring that local information be processed (e.g., Briand, 1994; Shedden & Reid, 2001). Furthermore, Dulaney and Marks (2007) showed that such global dominance can be eliminated. They found that extensive training at local identification eliminated interference from the global forms in the compound stimuli. Also, local interference was found after extensive training on local features. Thus, the usual nature of global/local processing can be modified by attentional manipulations. However, it took over 10,000 training trials to achieve this modification.

The global and local letter task (Navon, 1977) has also been used to prime global and local processing in other tasks. For example, it has been shown that priming subjects with global processing improved face recognition accuracy whereas priming with local processing impaired face recognition accuracy (Macrae & Lewis, 2002). On the other hand, a local superiority effect was demonstrated when subjects who had prior local processing were faster at face recognition in a facial composite task than were subjects who had prior global processing (Weston & Perfect, 2005).

The implication of these findings is that trainers need to keep in mind the degree to which local processing is required in a given task. When local processing is necessary, extensive training might need to be provided.

B. Stress conditions

Performance changes with level of stress on the trainee. At low levels of stress, performance might be poor, but as stress increases gradually, performance improves. At

a certain point, stress level is optimal for performance in any given task. Beyond the optimum, additional stress might degrade performance, and when stress becomes extreme the trainee might choke or panic (Staal, Bolton, Yaroush, & Bourne, 2008). However, stress has been shown to affect speed and accuracy of response differently. For example, the stress that comes from fatigue developed as a result of continuous work on a task leads to faster but less accurate performance (see the discussion above of speed-accuracy tradeoffs; Healy et al., 2004). Similarly, Wolfe, Horowitz, Cade, and Czeisler (2000) found that sleep deprivation led to an increase in errors on a visual search task for a target among varying numbers of distractors as well as to a reduction in the slope of the function relating response time to the number of distractors (see also Horowitz, Cade, Wolfe, & Czeisler, 2003). Thus, sleepy observers responded quickly but carelessly. Consequently, adding stressors to a training regime could be harmful (e.g., in the case of accuracy) or beneficial (e.g., when speed is the primary requirement) depending on what aspects of the task are most crucial and on the ambient level of stress. The implication of these findings for trainers is that they need to be aware of both trainee stress level and whether response speed or accuracy needs to be maximized.

C. Situational awareness

As automation has increased in many areas of life, the issue of how to maintain situational awareness (SA) has become crucial. SA is specific to dynamic systems in human-system interactions. High SA is generally required, but is not enough on its own, for high performance. SA involves not only an awareness of what is happening but also the implications for possible future outcomes (Endsley, 1995). Two things are necessary for maintaining SA: selective attention and long-term memory. Selective attention is needed to perceive or notice the important events in the situation, and long-term memory is needed to update knowledge of the situation. Most important is the trade-off between workload and SA (Wickens, 2002). As automation increases, workload decreases, but SA also decreases. The decrease in SA is due to both less monitoring of automated processes and less memory for the system state because changes in that state were not made by the human operator but by another agent (automation) (Endsley, 1995). The best way to mitigate this problem is still being researched (Wickens, 2008) (also see Dekker & Hollnagel, 2004; Dekker & Woods, 2002, for some criticisms of the concept of SA). In general, little is known at present concerning how to enhance SA through training, especially when automated systems are involved.

D. Just-in-time training

Learners need relevant task-specific information and skills to perform learning tasks and to learn from them. This necessary information must be active in working memory when performing the task. One way to reach this goal is to present the necessary information and skill training before the learners start working on the task, so that the knowledge and skills are encoded in schemas in long-term memory and subsequently activated in working memory if or when needed for the task ("just-in-case" training). Another way is to present the necessary information or skill training precisely when the learners need them during task performance. In this case, information and skill are

activated in working memory when they are necessary to perform the learning task. This method of training is called “just-in-time training” (JIT, JITT or JiT, also called “on-the-spot-training,” “on-call experts,” “real-time support,” “point-of-use information,” and “on-the-job” training). There is not an unequivocal answer to the question of which of the two ways (training before or just in time) is better. For tasks with a high-intrinsic complexity, it seems advisable to present the relevant information or skill training before the learners start on the learning tasks. Because learners have little cognitive capacity left for additional processing while working on the tasks, the simultaneous processing of intrinsically complex information or skills can easily lead to cognitive overload. If the information or skills are studied beforehand, a cognitive schema may be constructed in long-term memory that can subsequently be activated in working memory during task performance. Low-complexity information or skills, however, may better be presented precisely when learners need them during their work on the learning tasks. Because of their low-complexity, there is little or no chance of cognitive overload (Kester, Kirschner, & van Merriënboer, 2006; Kester, Kirschner, van Merriënboer, & Bäumer, 2001). Further research is necessary to confirm this speculation with unequivocal evidence as to when just-in-time training is desirable and superior to alternative training regimens.

VI. Summary and conclusions

This paper has reviewed the empirical and theoretical literature on training. This review strongly supports some training principles and more weakly supports other principles. These principles, even those that are strongly supported, do not necessarily apply for all tasks under all circumstances. Thus, it is important for a trainer to keep in mind certain distinctions that qualify these principles. Possibly the most critical of these distinctions is the difference between skill and knowledge (sometimes equated with the distinction between procedural and declarative information or the difference between implicit and explicit learning). Optimal training will differ depending on whether developing skill or acquiring knowledge is the primary goal.

The review also acknowledges the three fundamental cognitive processes underlying training, namely acquisition, retention, and transfer. Training principles in some cases apply differentially across those processes, such that some manipulations might facilitate acquisition but impede retention and/or transfer. Likewise, some training principles might impact particular performance measures but not others, especially under conditions involving a speed-accuracy tradeoff. Trainers need to be alert to the primary goal of training, which in some cases might be training efficiency but in other cases might be durability or generalizability. Similarly, trainers need to recognize the aspects of behavior that are most important to be optimized by training, which in some cases will be accuracy and in other cases speed of response.

Beyond the training principles that have been described, there are certain miscellaneous considerations about training that might impact how and when those principles are utilized. Among these is an assessment of the degree to which the task involves local versus global processing, keeping in mind that typically global processing takes precedence. Another consideration is the stress level induced by the training

context or brought to training by the trainee because it is well known that performance in general varies from poor to optimal as a function of stress level. Situational awareness is necessary for good performance in any training task or context, and so it should be promoted by the trainer. These last two considerations are related: Supra-optimal stress is known to shrink the perceptual field, thereby causing reduced situational awareness and the possibility of ignoring relevant information (Staal et al., 2008). The final consideration relates to when to provide task-relevant training. Typically, training is given well in advance of performance in the field. It is possible, however, that training of a part of a complex task might be effectively given only right before that part of the task is needed. The conditions under which such just-in-time training is effective are yet to be determined.

The training principles outlined here should be applicable in a variety of real-world training contexts including the training of astronauts and other military personnel. However, these are training principles, not training guidelines and certainly not training specifications (Salas et al., 1999). This review provides the first step in the design of optimal training programs. Additional developmental or applied research needs to be undertaken to translate these principles into guidelines and, subsequently, to specifications. Although this review focuses on training principles, it also offers brief suggested guidelines that might be examined and elaborated in the future. Particular applications must be based on research that refines the guidelines and translates them into usable training specifications.

References

Ackerman, P. L. (2007). New developments in understanding skilled performance. *Current Directions in Psychological Science 16*, 235–239.

Adams, J. A., & Hufford, L. E. (1962). Contributions of a part-task trainer to the learning and relearning of a time-shared flight maneuver. *Human Factors, 4*, 159-170.

Allen, G. A., Mahler, W. A., & Estes, W. K. (1969). Effects of recall tests on long-term retention of paired associates. *Journal of Verbal Learning and Verbal Behavior, 8*, 463-470.

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Anderson, L. S., Healy, A. F., Kole, J. A., & Bourne, L. E., Jr. (2010, April). *The clicker technique: An effective method of training compression*. Poster to be presented at the 80th Annual Convention of the Rocky Mountain Psychological Association, Denver, CO.

Anderson, R. C. (1968). Part-task versus whole-task procedures for teaching a problem-solving skill to first graders. *Journal of Educational Psychology 69*, 207-214.

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review, 105*, 442–481.

Au, T. K. (1990). Children's use of information in word learning. *Journal of Child Language*, *17*, 393-416.

Baddeley, A. (2007). *Working memory, thought, and action*. New York: Oxford University Press.

Bahrick, H. P. (1979). Maintenance of knowledge: Questions about memory we forgot to ask. *Journal of Experimental Psychology: General, 108*, 296-308.

Bahrick, H. P. (2005). The long-term neglect of long-term memory: Reasons & remedies. In A. F. Healy (Ed.), *Experimental cognitive psychology & its applications* (pp. 89-100). Washington, DC: American Psychological Association.

Bahrick, H. P., Bahrick, L. E., Bahrick, A. S., & Bahrick, P. E. (1993). Maintenance of foreign language vocabulary and the spacing effect. *Psychological Science, 4*, 316-321.

Bahrick, H. P., & Phelps, E. (1987). Retention of Spanish vocabulary over 8 years. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 344-349.

Barshi, I., & Healy, A. F. (1998). Misunderstandings in voice communication: Effects of fluency in a second language. In A. F. Healy & L. E. Bourne, Jr. (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 161-192). Mahwah, NJ: Erlbaum.

Barshi, I., & Healy, A. F. (2002). The effects of mental representation on performance in a navigation task. *Memory & Cognition*, *30,* 1189-1203.

Battig, W. F. (1972). Intratask interference as a source of facilitation in transfer and retention. In R. F. Thompson & J. F. Voss (Eds.), *Topics in learning and performance* (pp. 131-159). New York: Academic Press.

Battig, W. F. (1979). The flexibility of human memory. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory* (pp. 23-44). Hillsdale, NJ: Erlbaum.

Beier, M. E., & Ackerman, P. L. (2005). Age, ability and the role of prior knowledge on the acquisition of new domain knowledge. *Psychology and Aging, 20,* 341–355.

Beilock, S. L., Bertenthal, B. I., McCoy, A. M., & Carr, T. H. (2004). Haste does not always make waste: Expertise, direction of attention, and speed versus accuracy in performing sensorimotor skills. *Psychonomic Bulletin & Review, 11,* 373-379.

Bjork, E. L., deWinstanley, P. A., & Storm, B. C. (2007). Learning how to learn: Can the outcome of different strategies enhance subsequent encoding? *Psychonomic Bulletin & Review, 14*, 207-211.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185-205). Cambridge, MA: MIT Press.

Bjork, R. A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher & A. Koriat (Eds.), *Attention and performance XVII: Cognitive regulation of performance. Interaction of theory and application* (pp. 435-459). Cambridge, MA: MIT Press.

Bjork, R. A. (2001, March). How to succeed in college: Learn how to learn. *APS Observer, 14*, 9.

Bonk, W. J., & Healy, A. F. (2010). Learning and memory for sequences of pictures, words, and spatial locations: An exploration of serial position effects. *American Journal of Psychology*, *123*, 137-168.

Bourne, L. E., Jr., & Archer, E. J. (1956). Time continuously on target as a function of distribution of practice. *Journal of Experimental Psychology, 51,* 25-33.

Bourne, L. E., Jr., Dodd, D. H., Guy, D. E., & Justesen, D. R. (1968). Response-contingent intertrial intervals in concept identification. *Journal of Experimental Psychology, 76,* 601-608.

Bourne, L. E., Jr., Guy, D. E., Dodd, D. H., & Justesen, D. R. (1965). Concept identification: The effects of varying length and informational components of the intertrial interval. *Journal of Experimental Psychology, 69,* 624-629.

Bourne, L. E., Jr., Healy, A. F., Kole, J. A., & Graham, S. M. (2006). Strategy shifts in classification skill acquisition: Does memory retrieval dominate rule use? *Memory & Cognition, 34,* 903-913.

Bourne, L. E., Jr., Healy, A. F., Kole, J. A., & Raymond, W. D. (2004, November). *Bases for strategy choice and strategy transitions in binary classification tasks*. Paper presented at the 45th Annual Meeting of the Psychonomic Society, Minneapolis, MN.

Bourne, L. E., Jr., Healy, A. F., Parker, J. T., & Rickard, T. C. (1999). The strategic basis of performance in binary classification tasks: Strategy choices and strategy transitions. *Journal of Memory and Language, 41,* 223-252.

Bourne, L. E., Jr., Healy, A. F., Pauli, P., Parker, J. T., & Birbaumer, N. (2005). The influence of stimulus array on training of a speeded response. *American Journal of Psychology, 118,* 385-411.

Bourne, L. E., Jr., Parker, J. T., Healy, A. F., & Graham, S. M. (2000, November). *Strategy transitions and strategy retention in binary classification tasks*. Paper presented at the 41st annual meeting of the Psychonomic Society, New Orleans, LA.

Bourne, L. E., Jr., Raymond, W. D., & Healy, A. F. (2010). Strategy selection and use during classification skill acquisition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 500-514.

Briand, K. A. (1994). Selective attention to global and local structure of objects: Alternative measures of nontarget processing. *Perception & Psychophysics, 55,* 562–574.

Broadbent, D. E. (1975). The magic number seven after fifteen years. In A. Kennedy & A. Wilkes (Eds.), *Studies in long-term memory* (pp. 3–18). New York: Wiley.

Brown, N. R., & Siegler, R. S. (1996). Long-term benefits of seeding the knowledge base. *Psychonomic Bulletin & Review, 3,* 385-388.

Brown, N. R., & Siegler, R. S. (2001). Seeds aren't anchors. *Memory & Cognition, 29,* 405-412.

Brydges, R., Carnahan, H., Backstein, D., & Dubrowski, A. (2007). Application of motor learning principles to complex surgical tasks: Searching for the optimal practice schedule. *Journal of Motor Behavior 39*, 40-48.

Buck-Gengler, C. J., & Healy, A. F. (2001). Processes underlying long-term repetition priming in digit data entry. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 879-888.

Butler, A. C., Karpicke, J. D., & Roediger, H. L. III. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology: Applied*, *13*, 273-281.

Butler, A. C., & Roediger, H. L., III (2006, May). *Feedback neutralizes the detrimental effects of multiple-choice testing*. Poster presented at the annual meeting of the Association for Psychological Science, New York.

Carlson, R. A., & Yaure, R. G. (1990). Practice schedules and the use of component skills in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 484-496.

Carpenter, S. K., & DeLosh, E. L. (2005). Application of the testing and spacing effects to name learning. *Applied Cognitive Psychology, 19,* 619-636.

Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition, 34,* 268-276.

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition*, *20*, 633-642.

Clawson, D. M., Healy, A. F., Ericsson, K. A., & Bourne, L. E., Jr. (2001). Retention and transfer of Morse code reception skill by novices: Part-whole training. *Journal of Experimental Psychology: Applied, 7,* 129-142.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87-185.

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science, 19,* 51-57.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11,* 671-684.

Crutcher, R. J., & Healy, A. F. (1989). Cognitive operations and the generation effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 669-675.

Cull, W. L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology, 14,* 215-235.

Cull, W. L., Shaughnessy, J. J., & Zechmeister, E. B. (1996). Expanding understanding of the expanding-pattern-of-retrieval mnemonic: Toward confidence in applicability. *Journal of Experimental Psychology: Applied, 2,* 365-378.

Decety, J., Jeannerod, M., & Preblanc, C. (1989). The timing of mentally represented actions. *Behavioral Brain Research, 72,* 35-42.

Dekker, S., & Hollnagel, E. (2004). Human factors and folk models. *Cognition, Technology, and Work, 6,*79–86.

Dekker, S. W. A., & Woods, D. D. (2002). MABA-MABA or abracadabra? Progress on human-automation coordination. *Cognition, Technology, and Work, 4,* 240–244.

Delaney, P. F., Reder, L. M., Staszewski, J. J., & Ritter, F. E. (1998). The strategy-specific nature of improvement: The power law applies by strategy within task. *Psychological Science, 9,* 1-7.

Dempster, F. N. (1996). Distributing and managing the conditions of encoding and practice. In E. L. Bjork, & R. A. Bjork (Eds.), *Handbook of perception and cognition: Memory* (pp. 317–344). San Diego, CA: Academic Press.

deWinstanley, P. A., & Bjork, E. L. (2004). Processing strategies and the generation effect: Implications for making a better reader. *Memory & Cognition, 32,* 945-955.

Doane, S. M., Alderton, D. L., Sohn, Y.W., & Pellegrino, J. W. (1996). Acquisition and transfer of skilled performance: Are visual discrimination skills stimulus specific? *Journal of Experimental Psychology: Human Perception and Performance, 22,* 1218-1248.

Doane, S. M., Sohn, Y.W., & Schreiber, B. (1999). The role of processing strategies in the acquisition and transfer of a cognitive skill. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 1390-1410.

Driskell, J. E., Copper, C., & Moran, A. (1994). Does mental practice enhance performance? *Journal of Applied Psychology, 79,* 481-492.

Driskell, J. E., Willis, R. P., & Copper, C. (1992). Effect of overlearning on retention. *Journal of Applied Psychology, 77,* 615-622.

Dulaney, C. L., & Marks, W. (2007). The effects of training and transfer on global/local processing. *Acta Psychologica, 125,* 203-220.

Dutta, A., & Proctor, R.W. (1992). Persistence of stimulus-response compatibility effects with extended practice. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 18,* 801-809.

Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology*. New York: Teachers College, Columbia University. (Original work published 1885)

Eckman, F. R. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning, 27,* 315-330.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37,* 32–64.

Ericsson, K. A., Chase, W. G., & Faloon, S. (1980). Acquisition of a memory skill. *Science, 208,* 1181-1182.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100,* 363-406.

Fitts, P. M., & Deininger, R. L. (1954). S-R compatibility: Correspondence among paired elements within stimulus and response codes. *Journal of Experimental Psychology, 48,* 483-492.

Fitts, P. M., & Seeger, C. M. (1953). S-R compatibility: Spatial characteristics of stimulus and response codes. *Journal of Experimental Psychology, 46,* 199-210.

Fritz, C. O., Morris, P. E., Bjork, R. A., Gelman, R., & Wickens, T. D. (2000). When further learning fails: Stability and change following repeated presentations of text. *British Journal of Psychology, 91,* 493-511.

Gable, C. D., Shea, C. H., & Wright, D. L. (1991). Summary knowledge of results. *Research Quarterly for Exercise & Sport, 62,* 285-292.

Gardiner, J. M., Gardiner, M. M., & Gregg, V. H. (1983). The auditory recency advantage in longer term free recall is not enhanced by recalling prerecency items first. *Memory & Cognition, 11,* 616-620.

Gierut, J. A. (2001). Complexity in phonological treatment: Clinical factors. *Language, Speech, and Hearing Services in Schools, 32,* 229-241.

Glenberg, A. M. (1976). Monotonic and nonmonotonic lag effects in paired-associate and recognition memory paradigms. *Journal of Verbal Learning and Verbal Behavior, 15,* 1-16.

Goode, M. K., Geraci, L., & Roediger, H. L., III. (2008). Superiority of variable to repeated practice in transfer on anagram solution. *Psychonomic Bulletin & Review, 15*, 662-666.

Hardy, L., Mullen, R., & Jones, G. (1996). Knowledge and conscious control of motor actions under stress. *British Journal of Psychology, 87*, 621-636.

Haviland, S. E., & Clark, H. H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior, 13,* 512-521.

Healy, A. F. (2007). Transfer: Specificity and generality. In H. L. Roediger, III, Y. Dudai, & S. M. Fitzpatrick (Eds.), *Science of memory: Concepts* (pp. 271-275). New York: Oxford University Press.

Healy, A. F., & Bourne, L. E., Jr. (Eds.). (1995). *Learning and memory of knowledge and skills: Durability and specificity*. Thousand Oaks, CA: Sage.

Healy, A. F., Clawson, D. M., McNamara, D. S., Marmie, W. R., Schneider, V. I., Rickard, T. C., Crutcher, R. J., King, C., Ericsson, K. A., & Bourne, L. E., Jr. (1993). The long-term retention of knowledge and skills. In D. Medin (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 30, pp. 135-164). New York: Academic Press.

Healy, A. F., Fendrich, D. W., Crutcher, R. J., Wittman, W. T., Gesi, A. T., Ericsson, K. A., & Bourne, L. E., Jr. (1992). The long-term retention of skills. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes*, Volume 2 (pp. 87-118). Hillsdale, NJ: Erlbaum.

Healy, A. F., Kole, J. A., Buck-Gengler, C. J., & Bourne, L. E., Jr. (2004). Effects of prolonged work on data entry speed and accuracy. *Journal of Experimental Psychology: Applied, 10,* 188-199.

Healy, A. F., Shea, K. M., Kole, J. A., & Cunningham, T. F. (2008). Position distinctiveness, item familiarity, and presentation frequency affect reconstruction of order in immediate episodic memory. *Journal of Memory and Language, 58,* 746-764.

Healy, A. F., Wohldmann, E. L., & Bourne, L. E., Jr. (2005). The procedural reinstatement principle: Studies on training, retention, and transfer. In A. F. Healy (Ed.), *Experimental cognitive psychology and its applications* (pp. 59-71). Washington, DC: American Psychological Association.

Healy, A. F., Wohldmann, E. L., Kole, J. A., Schneider, V. I., Shea, K. M., & Bourne, L. E., Jr. (in press). Training for efficient, durable, and flexible performance in the military. In W. Arthur, Jr., E. A. Day, W. Bennett, Jr., & A. Portrey (Eds.), *Individual and team skill decay: State of the science and implications for practice*. New York: Taylor & Francis.

Healy, A. F., Wohldmann, E. L., Parker, J. T., & Bourne, L. E., Jr. (2005). Skill training, retention, and transfer: The effects of a concurrent secondary task. *Memory & Cognition*, *33*, 1457-1471.

Healy, A. F., Wohldmann, E. L., Sutton, E. M., & Bourne, L. E., Jr. (2006). Specificity effects in training and transfer of speeded responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 534-546.

Hintzman, D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola symposium* (pp. 77-99). Potomac, MD: Erlbaum.

Hockey, G. R. J., & Earle, F. (2006). Control over the scheduling of simulated office work reduces the impact of workload on mental fatigue and task performance. *Journal of Experimental Psychology: Applied, 12,* 50-65.

Holding, D. H. (1965). *Principles of training*. London: Pergamon.

Horowitz, T. S., Cade, B. E., Wolfe, J. M., & Czeisler, C. A. (2003). Searching night and day: A dissociation of effects of circadian phase and time awake on visual selective attention and vigilance. *Psychological Science, 14,* 549-557.

Howe, M. J. A. (1970). Repeated presentation and recall of meaningful prose. *Journal of Educational Psychology*, *61*, 214-219.

Hsiao, A. T., & Reber, A. S. (2001). The dual-task SRT procedure: Fine-tuning the timing. *Psychonomic Bulletin & Review, 8*, 336-342.

Ivancic, K., & Hesketh, B. (2000). Learning from errors in a driving simulation: Effects on driving skill and self-confidence. *Ergonomics, 43,* 1966-1984.

Izawa, C. (1992). Test trial contributions to optimization of learning processes: Study/test trials interactions. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes,* Volume 2 (pp. 1-33). Hillsdale, NJ: Erlbaum.

Kahana, M. J., & Howard, M. W. (2005). Spacing and lag effects in free recall of pure lists. *Psychonomic Bulletin & Review, 12,* 159-164.

Karpicke, J. D., & Roediger, H. L., III. (2010). Is expanding retrieval a superior method for learning text materials? *Memory & Cognition, 38,* 116-124.

Kay, H. (1955). Learning and retaining verbal material. *British Journal of Psychology, 46,* 81-100.

Kellogg, R. T., Friedman, A., Johnson, P., & Rickard, T. C. (2005). Domain-specific knowledge in intellectual skills: A symposium in honor of Lyle E. Bourne, Jr. In A. F. Healy (Ed.), *Experimental cognitive psychology and its applications* (pp. 3-15). Washington, DC: American Psychological Association.

Kern, R. S., Green, M. F., Mintz, J., & Liberman, R. P. (2003). Does 'errorless learning' compensate for neurocognitive impairments in the work rehabilitation of persons with schizophrenia? *Psychological Medicine, 33,* 433-442.

Kester, L., Kirschner, P. A., & van Merriënboer, J. J. G. (2006). Just-in-time information presentation: Improving learning a troubleshooting skill. *Contemporary Educational Psychology, 31,* 167-185.

Kester, L., Kirschner, P. A., van Merriënboer, J. J. G., & Bäumer, A. (2001). Just-in-time information presentation and the acquisition of complex cognitive skills. *Computers in Human Behavior, 17,* 373–391.

Kimchi, R. (1992). Primacy of wholistic processing and global/local paradigm: A critical review. *Psychological Bulletin, 112,* 24–38.

Kinchla, R. A. (1974). Detecting target elements in multi-element arrays: A confusability model. *Perception & Psychophysics, 15,* 149–158.

Kiran, S., & Thompson, C. K. (2003). The role of semantic complexity in treatment of naming deficits: Training semantic categories in fluent aphasia by controlling exemplar typicality. *Journal of Speech, Language, and Hearing Research, 46,* 773-787.

Kohl, R. M., & Roenker, D. L. (1983). Mechanism involvement during skill imagery. *Journal of Motor Behavior, 15,* 179-190.

Kole, J. A., & Healy, A. F. (2007). Using prior knowledge to minimize interference when learning large amounts of information. *Memory & Cognition, 35,* 124-137.

Kole, J. A., Healy, A. F., & Bourne, L. E., Jr. (2008). Cognitive complications moderate the speed-accuracy tradeoff in data entry: A cognitive antidote to inhibition. *Applied Cognitive Psychology, 22*, 917-937.

Kole, J. A., Healy, A. F., Fierman, D. M., & Bourne, L. E., Jr. (2010). Contextual memory and skill transfer in category search. *Memory & Cognition, 38,* 67-82.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 127*, 349-370.

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review, 14*, 219-224.

Kornell, N. & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science*, *19*, 585-592.

Kozhevnikov, M. (2007). Cognitive styles in the context of modern psychology: Toward an integrated framework of cognitive style. *Psychological Bulletin, 133,* 464-481.

Krueger, W. C. F. (1929). The effect of overlearning on retention. *Journal of Experimental Psychology, 12*, 71-78.

Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625-632). New York: Academic Press.

LaVoie, N. N., Bourne, L. E., Jr., & Healy, A. F. (2002). Memory seeding: Representations underlying quantitative estimations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 1137-1153.

Lee, T. D., & Magill, R. A. (1983). The locus of contextual interference in motor-skill acquisition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 730-746.

Lee, Y., & Vakoch, D. A. (1996). Transfer and retention of implicit and explicit learning. *British Journal of Psychology, 87,* 637-651.

Logan, G. D., & Klapp, S. T. (1991). Automatizing alphabet arithmetic: I. Is extended practice necessary to produce automaticity? *Journal of Experimental Psychology: Learning, memory, & Cognition, 17,* 179-195.

Macrae, C. N., & Lewis, H. L. (2002). Do I know you? Processing orientation and face recognition. *Psychological Science, 13,* 194-196.

Marmie, W. R., & Healy, A. F. (1995). The long-term retention of a complex skill. In A. F. Healy & L. E. Bourne, Jr. (Eds.), *Learning and memory of knowledge and skills*: *Durability and specificity* (pp. 30-65). Thousand Oaks, CA: Sage.

Marsh, E. J., Roediger, H. L., III, Bjork, R. A., & Bjork, E. L. (2007). The memorial consequences of multiple-choice testing. *Psychonomic Bulletin & Review, 14*, 194-199.

Marshall, L., & Born, J. (2007). The contribution of sleep to hippocampus- dependent memory consolidation. *Trends in Cognitive Sciences, 11*, 442– 450.

Martin, M. (1979). Local and global processing: The role of sparsity. *Memory & Cognition, 7*, 476–484.

Massa, L. J., & Mayer, R. E. (2006). Testing the ATI hypothesis: Should multimedia instruction accommodate verbalizer-visualizer cognitive style? *Learning and Individual Differences, 16*, 321-336.

Masters, R. S. W. (1992). Knowledge, knerves and know-how: the role of explicit versus implicit knowledge in the breakdown of a complex motor skill under pressure. *British Journal of Psychology, 83*, 343-358.

Maxwell, J. P., Masters, R. S. W., Kerr, E., & Weedon, E. (2001). The implicit benefit of learning without errors. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 54A*, 1049-1068.

Mayer, R. E. (2001). *Multimedia learning*. New York: Cambridge University Press.

Mayer, R. E., Stull, A., DeLeeuw, K., Almeroth, K., Bimber, B., Chun, D., Bulger, M., Campbell, J., Knight, A., & Zhang, H. (2008). Clickers in college classrooms: Fostering learning with questioning methods in large lecture classes. *Contemporary Educational Psychology*, *34*, 51-57.

Mazzoni, G., & Nelson, T. O. (Eds.) (1998). *Metacognition and cognitive neuropsychology: Monitoring and control processes*. Mahwah, NJ: Lawrence Erlbuam Associates.

McDaniel, M. A., & Butler, A. C. (in press). A contextual framework for understanding when difficulties are desirable. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: A festschrift in honor of Robert A. Bjork.* New York: Psychology Press.

McDaniel, M. A., & Einstein, G. O. (2005). Material appropriate difficulty: A framework for determining when difficulty is desirable for improving learning. In A. F. Healy (Ed.), *Experimental cognitive psychology and its applications* (pp. 73-85). Washington, DC: American Psychological Association.

McDaniel, M. A., & Fisher, R. P. (1991). Tests and test feedback as learning sources. *Contemporary Educational Psychology, 16*, 192-201.

McDaniel, M. A., Roediger, H. L., III., & McDermott, K. B. (2007). Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin & Review, 14,* 200-206.

McDaniel, M. A., & Schlager, M. S. (1990). Discovery learning and transfer of problem-solving skills. *Cognition and Instruction, 7,* 129-159.

McNamara D. S. (2001). Reading both high-coherence and low-coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology, 55*, 51-62.

McNamara, D. S., & Healy, A. F. (1995). A procedural explanation of the generation effect: The use of an operand retrieval strategy for multiplication and addition problems. *Journal of Memory and Language, 34,* 399-416.

McNamara, D. S., & Healy, A. F. (2000). A procedural explanation of the generation effect for simple and difficult multiplication problems and answers. *Journal of Memory and Language, 43,* 652-679.

McNamara, D. S. & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes, 22*, 247-288.

McNamara, D.S., Kintsch, E., Songer N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from texts. *Cognition and Instruction, 14,* 1-43.

McNevin, N. H., Shea, C. H., & Wulf, G. (2003). Increasing the distance of an external focus of attention enhances learning. *Psychological Research, 67,* 22-29.

Means, M., & Voss, J. (1985). Star Wars: A developmental study of expert and novice knowledge structures. *Memory and Language, 24*, 746-757.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81-97.

Minas, S. C. (1978). Mental practice of a complex perceptual motor skill. *Journal of Human Movement Studies, 4*, 102-109.

Moravcsik, J. E., & Kintsch, W. (1993). Writing quality, reading skills, and domain knowledge as factors in text comprehension. *Canadian Journal of Experimental Psychology, 47,* 360-374.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16*, 519-533.

Morris, P. E., & Fritz, C. O. (2000). The name game: Using retrieval practice to improve the learning of names. *Journal of Experimental Psychology: Applied, 6*, 124-129.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology, 9*, 353–383.

Navon, D. (1991). Testing a queue hypothesis for the processing of global and local information. *Journal of Experimental Psychology: General, 120*, 173–189.

Navon, D., & Norman, J. (1983). Does global precedence really depend on visual angle. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 955–965

Naylor, J. C., & Briggs G. E. (1963). Effects of task complexity and task organization on the relative efficiency of part and whole training methods. *Journal of Experimental Psychology, 65*, 217-224.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the power law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1-55). Hillsdale, NJ: Erlbaum.

Newell, K. M., Carlton, M. J., Fisher, A.T., & Rutter, B. G. (1989). Whole-part training strategies for learning the response dynamics of microprocessor driven simulators. *Acta Psychologica, 71*, 197-216.

Nipher, F. E. (1878). On the distribution of errors in numbers written from memory. *Transactions of the Academy of Science of St. Louis, 3*, ccx - ccxi.

Norman, G. (2004). Editorial -- What's the active ingredient in active learning? *Advances in Health Sciences Education, 9*, 1-3.

Osgood, C. E. (1949). The similarity paradox in human learning: A resolution. *Psychological Review, 56*, 132-143.

Pachella, R. G. (1974). The interpretation of reaction time in information-processing research. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition* (pp. 41-82). Hillsdale, NJ: Erlbaum.

Paquet, L. (1992). Global and local processing in nonattended objects: A failure to induce local processing dominance. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 512–529.

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 3-8.

Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2009). Learning styles: Concepts and evidence. *Psychological Science in the Public Interest, 9,* 105-119.

Pashler , H., Rohrer, D., Cepeda, N. J., & Carpenter, S. K. (2007). Enhancing learning and retarding forgetting: Choices and consequences. *Psychonomic Bulletin & Review, 14*, 187-193.

Peck A. C., & Detweiler, M. C. (2000). Training concurrent multistep procedural tasks. *Human Factors, 42*, 379-389.

Pellegrino, J. W., Doane, S. M., Fischer, S. C. & Alderton, D. (1991). Stimulus complexity effects in visual comparisons: The effects of practice and learning context. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 781-791.

Penney, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition, 17,* 398-422.

Perels, F., Gürtler, T., & Schmitz, B. (2005). Training of self-regulatory and problem-solving competence. *Learning and Instruction, 15*, 123-139.

Péruch, P., & Wilson, P. N. (2004). Active versus passive learning and testing in a complex outside built environment. *Cognitive Processing, 5,* 218-227.

Postman, L., & Underwood, B. J. (1973). Critical issues in interference theory. *Memory & Cognition, 1,* 19-40.

Proctor, R. W., & Dutta, A. (1995). *Skill acquisition and human performance*. Thousand Oaks, CA: Sage publications.

Proctor, R.W., & Vu, K.-P.L. (2006). *Stimulus-response compatibility principles: Data, theory, and application*. Boca Raton, FL: CRC Press.

Rickard, T. C. (1997). Bending the power law: A CMPL theory of strategy shifts and the automatization of cognitive skills. *Journal of Experimental Psychology: General, 126,* 288-311.

Rickard, T. C. (2004). Strategy execution in cognitive skill learning: An item-level test of candidate models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 65-82.

Rickard, T. C., Cai, D. J., Reith, C. A., Jones, J., &Ard, M. C. (2008). Sleep does not

enhance motor sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 834-842.

Rickard, T. C., Lau, J. S.-H., & Pashler, H. (2008). Spacing and the transition from calculation to retrieval. *Psychonomic Bulletin & Review 15*, 656-661.

Robertson, L. C., Egly, R., Lamb, M. R., & Kerth, L. (1993). Spatial attention and cuing to global and local levels of hierarchical structure. *Journal of Experimental Psychology: Human Perception and Performance, 19*, 471–487.

Roediger, H. L., III. (2008). Relativity of remembering: Why the laws of memory vanished. *Annual Review of Psychology, 59*, 225-254.

Roediger, H. L., III. (2009, November). *The critical role of retrieval in enhancing long-term memory: From the laboratory to the classroom*. Keynote address at the 50th Annual Meeting of the Psychonomic Society, Boston, MA.

Roediger, H. L., III, Weldon, M. S., & Challis, B. H. (1989). Explaining dissociations between implicit and explicit measures of retention: A processing account. In H. L. Roediger, III, & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 3-41). Hillsdale, NJ: Erlbaum.

Rohrer, D., & Taylor, K. (2006).The effects of overlearning and distributed practice on the retention of mathematics knowledge. *Applied Cognitive Psychology, 20*, 1209-1224.

Rohrer, D., Taylor, K., Pashler, H., Wixted, J. T., & Cepeda, N. J. (2005). The effect of overlearning on long-term retention. *Applied Cognitive Psychology, 19*, 361-374.

Rubin, D. C., & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review, 103*, 734-760.

Salas, E., Cannon-Bowers, J. A., & Blickensderfer, E. L. (1999). Enhancing reciprocity between training theory and practice: Principles, guidelines, and specifications. In G. R. Ferris (Ed.), *Research in personnel and human resources management* (Vol. 17, pp. 291-321). Stamford, CT: JAI Press.

Schmidt, R. A. (1975). A schema theory of discrete motor skill learning. *Psychological Review, 82*, 225-260.

Schmidt, R. A. (1991). Frequent augmented feedback can degrade learning: Evidence and interpretations. In J. Requin & G. E. Stelmach (Eds.), *Tutorials in motor neuroscience* (pp. 59-75). Dordrecht, The Netherlands: Kluwer.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3,* 207-217.

Schmidt R. A., & Lee, T. D. (1999). *Motor control and learning: A behavioral emphasis.* Champaign, IL: Human Kinetics.

Schmidt, R. A., Young, D. E., Swinnen, S., & Shapiro, D. C. (1989). Summary knowledge of results for skill acquisition: Support for the guidance hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 352-359.

Schneider, V. I., Healy, A. F., & Barshi, I. (2004). Effects of instruction modality and readback on accuracy in following navigation commands. *Journal of Experimental Psychology: Applied, 10,* 245-257.

Schneider, V. I., Healy, A. F., & Bourne, L. E., Jr. (1998). Contextual interference effects in foreign language vocabulary acquisition and retention. In A. F. Healy & L. E. Bourne, Jr. (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 77-90). Mahwah, NJ: Erlbaum.

Schneider, V. I., Healy, A. F., & Bourne, L. E., Jr. (2002). What is learned under difficult conditions is hard to forget: Contextual interference effects in foreign vocabulary acquisition, retention, and transfer. *Journal of Memory and Language*, *46*, 419-440.

Schneider, V. I., Healy, A. F., Buck-Gengler, C. J., Barshi, I., & Bourne, L. E., Jr. (2007, July). *The effects of feedback on learning to follow navigation instructions*. Poster presented at the joint meeting of the Experimental Psychology Society and the Psychonomic Society, Edinburgh, Scotland.

Schneider, V. I., Healy, A. F., Ericsson, K. A., & Bourne, L. E., Jr. (1995). The effects of contextual interference on the acquisition and retention of logical rules. In A. F. Healy & L. E. Bourne, Jr. (Eds.), *Learning and memory of knowledge and skills: Durability and specificity* (pp. 95-131). Thousand Oaks, CA: Sage.

Schunk, D. H. (2005). Commentary on self-regulation in school contexts. *Learning and Instruction, 15*, 173-177.

Shanks, D. R., & Cameron, A. (2000). The effect of mental practice on performance in a sequential reaction time task. *Journal of Motor Behavior, 32,* 305-313.

Shea, C. H., & Wulf, G. (1999). Enhancing motor learning through external-focus instructions and feedback. *Human Movement Science, 18*, 553-571.

Shea, J. B., & Morgan, R. L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 179-187.

Shedden, J. M., & Reid, G. S. (2001). A variable mapping task produces symmetrical interference between global information and local information. *Perception & Psychophysics, 63*, 241–252.

Shepard R. N. (1987). Toward a universal law of generalization for psychological science. *Science, 237,* 1317-1323.

Simon, D., & Bjork, R. A. (2001). Metacognition in motor learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 907-912.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 592-604.

Smith, J. D., Redford, J. S., Washburn, D. A., & Taglialatela, L. A. (2005). Specific-token effects in screening tasks: Possible implications for aviation safety. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 1171-1185.

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review, 8,* 203-220.

Smith, T. A., & Kimball, D. R. (2010). Learning from feedback: Spacing and the delay-retention effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 80-95.

Sperling, R. A., Howard, B. C., Staley, R., & DuBois, N. (2004). Metacognition and self-regulated learning constructs. *Educational Research and Evaluation, 10,* 117-139.

Spiering, B. J., & Ashby, F. G. (2008). Initial training with difficult items facilitates information integration, but not rule-based category learning. *Psychological Science, 19*, 1169-1177.

Staal, M. A., Bolton, A. E., Yaroush, R. A., & Bourne, L. E., Jr. (2008). Cognitive performance and resilience to stress. In B. J. Lukey & V. Tepe (Eds.), *Biobehavioral resilience to stress* (pp. 259-300). Boca Raton, FL: CRC Press.

Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature, 437*, 1272–1278.

Storm, B. C., Bjork, R. A., & Storm, J. C. (2010). Optimizing retrieval as a learning event: When and why expanding retrieval practice enhances long-term retention. *Memory & Cognition, 38,* 244-253.

Terrace, H. S. (1963). Discrimination learning with and without "errors." *Journal of the Experimental Analysis of Behavior, 6,* 1-27.

Thompson, C. K., Shapiro, L. P., Tait, M. E., Jacobs, B. J., & Schneider, S. L. (1996). Training wh-question production in agrammatic aphasia: Analysis of argument and adjunct movement. *Brain and Language, 52,* 175–228.

Touron, D. R., Hoyer, W. J., & Cerella, J. (2004). Cognitive skill learning: Age-related differences in strategy shifts and speed of component operations. *Psychology and Aging, 19,* 567-580.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80,* 352-373.

Vakil, E., Hoffman, Y., & Myzliek, D. (1998). Active versus passive procedural learning in older and younger adults. *Neuropsychological Rehabilitation, 8,* 31-41.

Van Overschelde, J. P., & Healy, A. F. (2001). Learning of nondomain facts in high- and low-knowledge domains. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 1160-1171.

Underwood, B. J., & Ekstrand, B. R. (1967). Studies of distributed practice: XXIV. Differentiation and proactive inhibition. *Journal of Experimental Psychology, 74,* 574-580.

Walker, M. P. (2005). A refined model of sleep and the time course of memory formation. *Behavioral and Brain Sciences, 28,* 51–104.

Walker, M. P., & Stickgold, R. (2004). Sleep-dependent learning and memory consolidation. *Neuron, 44,* 121–133.

Walker, M. P., & Stickgold, R. (2006). Sleep, memory, and plasticity. *Annual Review of Psychology, 57,* 139 –166.

Ward, L. M. (1982). Determinants of attention to local and global features of visual forms. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 562–581.

Weston, N. J., & Perfect, T. J. (2005). Effects of processing bias on the recognition of composite face halves. *Psychonomic Bulletin & Review, 12,* 1038-1042.

Wickelgren, W.A. (1974). Single-trace fragility theory of memory dynamics. *Memory & Cognition, 2,* 775–780.

Wickens, C. D. (2002). Situation awareness and workload in aviation. *Current Directions in Psychological Science, 11,*128–133.

Wickens, C. D. (2008). Situation awareness: Review of Mica Endsley's 1995 articles on situation awareness theory and measurement. *Human Factors*, *50,* 397-403.

Wightman, D. & Lintern, G. (1985). Part-task training for tracking and manual control. *Human Factors*, *27*, 267-283.

Wixted, J. T. (2004). The psychology and neuroscience of forgetting. *Annual Review of Psychology, 55*, 235–269.

Wixted, J. T., & Carpenter, S. K. (2007). The Wickelgren power law and the Ebbinghaus savings function. *Psychological Science, 18,* 133-134.

Wohldmann, E. L., Healy, A. F., & Bourne, L. E., Jr. (2007). Pushing the limits of imagination: Mental practice for learning sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 254-261.

Wohldmann, E. L., Healy, A. F., & Bourne, L. E., Jr. (2008a). A mental practice superiority effect: Less retroactive interference and more transfer than physical practice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 823-833.

Wohldmann, E. L., Healy, A. F., & Bourne, L. E., Jr. (2008b). Global inhibition and midcourse corrections in speeded aiming. *Memory & Cognition, 36*, 1228-1235.

Wolfe, J. M., Horowitz, T. S., Cade, B. E., & Czeisler, C. A. (2000, November). *Searching when sleepy: Trading accuracy for efficiency*. Paper presented at the 41st annual meeting of the Psychonomic Society, New Orleans, LA.

Wolfe, M. B. W., Schreiner, M. E., Rehder, B., Laham, D., Foltz, P. W., Kintsch, W. & Landauer, T. K. (1998). Learning from text: Matching readers and text by latent semantic analysis. *Discourse Processes*, *25,* 309-336.

Wulf, G., McNevin, N. H., & Shea, C. H. (2001). The automaticity of complex motor skill learning as a function of attentional focus. *Quarterly Journal of Experimental Psychology, 54A,* 1143-1154.

Wulf, G., & Schmidt, R. A. (1997). Variability of practice and implicit motor learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 987-1006.

Wulf, G., & Shea, C. H. (2002). Principles derived from the study of simple skills do not generalize to complex skill learning. *Psychonomic Bulletin & Review, 9,* 185-211.

Wulf, G., Shea, C. H., & Whitacre, C. A. (1998). Physical-guidance benefits in learning a complex motor skill, *Journal of Motor Behavior, 30,* 367-380.

Young, M. D., Healy, A. F., Gonzalez, C., Dutt, V., & Bourne, L. E., Jr. (in press). Effects of training with added difficulties on RADAR detection. *Applied Cognitive Psychology*.

Acquisition and Transfer of Basic Skill Components

Robert W. Proctor, Motonori Yamaguchi, James D. Miles
Purdue University

Acquisition and Transfer of Basic Skill Components

Robert W. Proctor, Motonori Yamaguchi, James D. Miles
Purdue University

The goal of our part of the Training MURI was to study in detail basic tasks that isolate the perceptual, cognitive, and motor components of skill, examining factors that influence acquisition and transfer of these components. Speed and accuracy of response selection is a fundamental skill underlying performance of most tasks that is acquired rapidly with practice and can be studied effectively in the laboratory. Consequently, we focused on a detailed analysis of skill components in tasks that emphasize speeded response selection. The results of these studies support several fundamental principles of training, which we summarize in this report.

1. Introduction. Our research had the goal of examining factors that influence the acquisition and transfer of fundamental components of skill. For much of this research, we utilized the power of basic choice reaction tasks to isolate fundamental cognitive processes and allow rapid acquisition of skill within a single experimental session. The methods we used relied heavily, though not exclusively, on variants of spatial stimulus-response compatibility (SRC) tasks. The concept of SRC and the first investigations of compatibility effects are attributed to Paul M. Fitts (Fitts & Deininger, 1954; Fitts & Seeger, 1953), who founded the Psychology Branch of the Aero Medical Laboratory of the U.S. Army at Wright Field at the end of World War II. Perhaps more than anyone, he recognized the value of basic laboratory tasks for understanding processes involved in much more complex military tasks. This value has also been appreciated by other researchers associated with the military who have used SRC tasks in the investigation of human performance issues, including Earl A. Alluisi (Alluisi & Warm, 1990), Chief Scientist at the Air Force Human Resources Laboratory at Brooks Air Force Base in the first half of the 1980s and then Assistant for Training and Personnel Systems Technology in the Office of the Secretary of Defense in the last half of the 1980s. Thus, our work follows in a long tradition of exploiting the properties of SRC tasks to investigate a range of issues in human skilled performance, in this case, ones concerning practice and transfer effects.

For much of our research, we used two-choice reaction tasks. In the prototypical task, a stimulus can appear in a left or right location, and the performer is to press an assigned left or right response key as quickly as possible. Responses are on average about 50 ms faster when the task is performed with a compatible mapping of “press the left key to the left light and right key to the right light” than with an incompatible mapping of “press the left key to the right light and the right key to the left light.” Although performance improves with practice, this *SRC effect* remains evident even after relatively large amounts of practice (Dutta & Proctor, 1992; Fitts & Seeger, 1953).

We also used a variant of the task that has come to be known as the Simon task, after J. R. Simon (1990). For a Simon task, the relevant stimulus dimension is not the location of the stimulus but some non-spatial feature such as its color (often, red or

green). The *Simon effect* refers to the fact that responses are still faster, and often more accurate, when stimulus and response locations correspond than when they do not, even though stimulus location is defined as irrelevant to the task. The Simon effect has attracted considerable research interest in recent years because it enables investigation of how response selection is affected by features of a task that are not an explicit part of the instructed task goals. The Simon effect is typically attributed to long-term associations, or links, between particular stimuli and responses (e.g., left stimulus locations and left responses) that have been acquired through years of practice. Activation of the corresponding response is often described as occurring automatically by way of these long-term links when the appropriate stimulus occurs.

The research that we performed for the MURI had three parts: (a) transfer of newly acquired associations, (b) training with mixed mappings and tasks, and (c) performance of multiple tasks. In the following sections, we describe our main findings in these areas and implications of those findings for skills training.

2. Factors affecting transfer of learning. Our studies of transfer of learning used the following basic paradigm: In a practice session, subjects performed a two-choice spatial SRC task with an incompatible mapping (e.g., press "left" key to a stimulus that appears on the "right"; *incompatible-mapping task*). Then, in a transfer session, the subjects performed a Simon task in which they responded to a nonspatial stimulus attribute (e.g., color). Thus, the spatial relation between stimulus and response in the practice task was task-relevant, but it became task-irrelevant in the transfer task. The logic behind the research is that practice establishes new links between the stimuli and their assigned responses (sometimes called short-term links) that, in the case of an incompatible mapping, are counter to the long-term links that produce the typical Simon effect. After performing the incompatible-mapping task, the advantage for the spatially corresponding responses in the Simon task is eliminated and in some cases reversed (Proctor & Lu, 1999). This outcome implies that the incompatible stimulus-response (S-R) links acquired for the practice task are transferred to a subsequent task even though they are no longer relevant. This experimental paradigm is particularly well suited to investigating factors that affect transfer of learning because of the many manipulations of sensory modalities, modes for presenting location information, response modes, and so on, that can be made for the practice and transfer tasks.

Perhaps the most striking outcome of the practice/transfer tasks is how easy it is to overcome or counteract effects of long-term associations between stimuli and responses. The benefit for spatial correspondence is eliminated by less than 100 trials of practice with an incompatible spatial mapping, and this elimination is equally apparent 5 minutes later, one day later, and a week later (Vu, Proctor, & Urcuioli, 2003; Tagliabue, Zorzi, Umiltà, & Bassignani, 2000). In other words, this small amount of training is sufficient to produce durable new S-R links that will override the pre-existing habitual response tendencies. With larger amounts of practice, the transfer task shows reversal of the Simon effect to favor the practiced incompatible S-R relation (Proctor & Lu, 1999), and shows a broader range of transfer (e.g., Proctor, Yamaguchi, & Vu, 2007). Transfer of the practice mapping occurs for auditory stimuli as well as visual stimuli, for arrow directions and spatial words, as well as physical locations, for various response modes (e.g., unimanual joystick movements, keypresses, as well as vocal utterances), and for

vertically oriented S-R sets as well as for horizontally oriented ones. Reversal of the Simon effect also occurs when trials with a task using a spatially incompatible mapping are intermixed with trials of the Simon task (e.g., Proctor, Vu, & Marble, 2003; see next section), a result also thought to reflect transfer of the task-defined S-R location links to the Simon trials, for which stimulus location is not relevant. Finally, the typical advantage for the corresponding location can also be offset simply by giving implementation instructions, in which instructions describe a specific goal of making a particular response quickly whenever a specific stimulus condition occurs (e.g., if a red stimulus appears in the left location, press the right key; Cohen, Bayer, Jaudas, & Gollwitzer, 2008; Miles & Proctor, 2008).

Many of the findings we have obtained with the practice/transfer paradigm can be accommodated within the quantitative framework developed by the MURI team, in which the strength of learned knowledge is represented by an activation function:

$$a_n = \sum_{i=1}^{n} \beta_i t_i^{-\lambda} \exp\{\alpha S_i\}, \tag{1}$$

where a_n represents the activation of target knowledge after n practice trials. Provided t_i, $\beta_i > 0$, a_n increases as n increases. The equation embraces a kind of *strength theory* that states that remembering is a function of the strength of the memory trace (representation) [but see Logan (1988) for a possible interpretation of the equation based on an instance theory]. Though the strength theory was originally proposed for learning of "declarative knowledge" (memory of facts), our experiments suggest that the model is also applicable to "procedural memory" (memory of acts). Furthermore, the experimental results imply that the strength of procedural memory is a function of practice amount so that extended practice can overcome the pre-existing habitual response tendencies to the environment. The fact that a greater amount of practice is needed in some conditions (e.g., for word stimuli) can be modeled in the framework by the learning rate β_i.

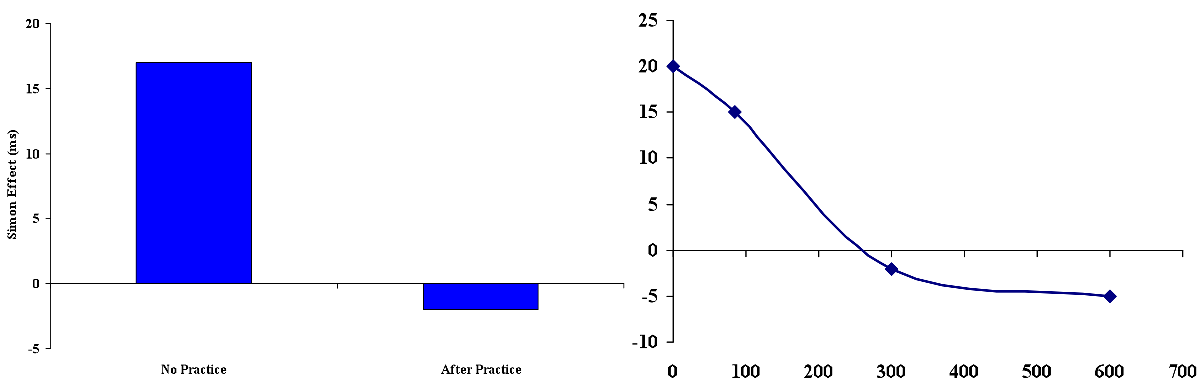


Figure 1. *Simon effect (a) with no prior practice and after < 100 trials of practice with an incompatible spatial mapping and (b) as a function of practice (Proctor, Yamaguchi, Zhang, & Vu, 2009).*

In the framework, efficiency of training is determined by number of trials (N), learning rate (β), contextual similarity (S), and time passage (t and λ). As noted, practice with an incompatible mapping increases the associative strength for the incompatible S-R link through increase in N. The strength of the incompatible S-R link is reflected in

reduction of the Simon effect, as in Figure 1a which shows that the Simon effect became smaller (or eliminated) after practice with the incompatible mapping. The relation between strength of S-R link and the amount of practice follows a power function. Consequently, when plotted against the number of practice trials (Figure 1b), the Simon effect initially decreases rapidly, but the amount of change deaccelerates over trials, eventually reaching an asymptote.

The learning rate β may be dependent on several factors such as learners' motivation and comprehensive capability, the effectiveness of instructions, time scheduling of training, and the difficulty of learning materials. In our experiments, we examined the transfer effect for different types of spatial stimuli (physical location of a circle, pointing direction of an arrow, the meaning of spatial words) and observed that the learning rate depended on this factor (Proctor et al., 2009). In particular, the transfer effect was evident after less than 100 trials of practice when the spatial information was conveyed by the physical location or the pointing direction of arrows. Although the Simon effect tended to be larger for the arrow stimuli than for the location stimuli, the size of the transfer effect was equivalent for the two types of stimuli. In contrast, after practice with the word stimuli for less than 100 trials, there was little indication of the transfer effect. Nevertheless, when the number of practice trials was increased to 300 trials, the transfer effect was observed (as shown in Figure 1b), which was as large as that for the location and arrow stimuli. Because responses were made by pressing the left and right keys, *set-level compatibility* (cf., Proctor & Wang, 1997) was higher for the location and arrow stimuli than for the word stimuli. Therefore, we conducted a similar experiment with vocal responses (i.e., saying "left" or "right"), for which set-level compatibility should be higher for the word stimuli than for the location and arrow stimuli. However, we found that the transfer effect was evident for the location stimuli after less than 100 practice trials, but it appeared for the word stimuli only after the number of trials was doubled, suggesting that the learning rate is not dependent on set-level compatibility but is determined by the stimulus type.

Another important aspect of transfer of learning is its limitations. According to the framework, learning is utilized better in a context that is similar to the original context in which the learning has taken place, the *principle of transfer specificity* (see Healy, Schneider, & Bourne's report). The influence of contextual similarity of the current trial to past trials is expressed by the exponential component of Equation 1, where S_i is the similarity of the ith practice trial to the current trial.

A well-known non-metric theory of similarity judgment is Tversky's (1977) *contrast model* in which an object or event is considered to be a set of unique features. Then, the similarity between two objects X_i and X_j is expressed by

$$S_{ij} = f(X_i \cap X_j) - g(X_i / X_j) - h(X_j / X_i). \tag{3}$$

A special case of the contrast model is the *feature overlap account* of contextual similarity (see Figure 2) in which the similarity between two task contexts (practice context C_p and test context C_t) is considered to be a function of the number of overlapping features between the contexts

$$S(C_p, C_t) = f(C_p \cap C_t), \tag{4a}$$

or more specifically,

$$S(C_p, C_t) = \sum_{i,j} M(x_i, y_j), \tag{4b}$$

where $x_i \in C_p$ and $y_j \in C_t$ with M being a matching function defined by $M(x_i, y_j) = 1$ if $x_i = y_j$ and $M(x_i, y_j) = 0$ if $x_i \neq y_j$.

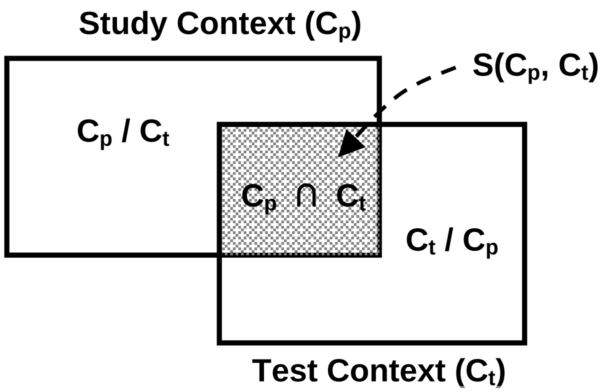


Figure 2. *A feature overlap account of contextual similarity.*

In our transfer studies, we examined boundary conditions of transfer of newly acquired associations by varying contextual features of the practice and transfer tasks. The results were consistent with the feature overlap account. For instance, the transfer effect is larger when the stimulus modalities (visual or auditory) match between the practice and transfer conditions than when they mismatch (Proctor et al., 2007; Vu et al., 2003); when the types of stimulus mode (location word, arrow direction, or physical location) match than when they mismatch (Proctor et al., 2009); when the response modes match than when they mismatch (Yamaguchi & Proctor, 2009); and when the stimuli and responses are oriented along the same spatial dimension (e.g., both horizontal) than along orthogonal dimensions (one vertical, the other horizontal; Vu, 2007; Proctor et al., 2007). Hence, transfer of newly acquired associations depends on overlap of contextual features present during practice and test.

According to the framework, influence of time passage (t and λ) is thought to be loss of learning; that is, learned skills dissipate over time if the skills are not used. However, there is a long debate in psychology as to whether dissipation of learning (or memory) is due to passive decay or interference. Depending on the theoretical position in this debate, one can formulate different models of skill dissipation. In our previous studies (Proctor et al., 2003), the transfer effect was as large a week after the practice session took place as 5 min. after the session. This finding suggests that learned S-R links did not decay even if participants did not perform the incompatible-mapping task for a week. On the contrary, we found that the transfer effect was essentially eliminated if there were intervening trials for which participants performed the incompatible-mapping task but with a different type of stimuli. In particular, participants were first provided with a practice session with word stimuli. Then, they performed another practice session with arrow stimuli. Finally, they transferred to the Simon task with the word stimuli. The Simon effect was larger than the effect observed for the group who was provided only with the first practice session (no intervening session) but as large as the control group who were not provided with the practice sessions. These results imply that the

intervening task “cut off” the learned incompatible S-R links. Hence, our results support interference as the cause of skill dissipation. Given that the MURI framework currently lacks specification of the mechanism that underlying skill dissipation, the framework can be further elaborated by incorporating a component that expresses interference of learning by intervening tasks.

3. Training with mixed mappings and tasks. People often have to be prepared to perform multiple tasks, any one of which must be performed when an appropriate event occurs, rather than performing a single task in isolation. Thus, it is important to know how performance of one task is influenced by the presence of other tasks to perform. We have investigated the influence of mixing compatible and incompatible mappings on choice-reaction tasks (Vu & Proctor, 2004; Yamaguchi & Proctor, 2006) and found that the performance advantage of the compatible mapping over the incompatible mapping is reduced or eliminated under mixed conditions. This finding can be attributed to subjects’ having to be prepared to perform the incompatible-mapping task at any moment during the session, so that they suppress the natural tendency to respond with a spatially compatible response to a stimulus. The advantage for the compatible spatial mapping is also lost when trials for which stimulus location is relevant (with only a single mapping) are mixed with Simon-task trials for which stimulus location is irrelevant (Proctor & Vu, 2002; Proctor et al., 2003). Also, the Simon effect increases somewhat when the spatial mapping for the location-relevant trials is compatible but reverses to favor the non-corresponding response when that mapping is incompatible.

We have examined the specificity of these mixing effects on performance in recent studies. Proctor and Vu (2009c) showed that the effects of task mixing on the spatial compatibility and Simon effects were reduced when the location information was presented in different modes (physical locations vs. location words) for the two tasks. In contrast, the mode distinction had little influence on the effects of mixing compatible and incompatible location mappings. These results imply that when location is relevant for one task and color for the other, the task-defined associations of locations to responses are mode specific, but when location is relevant for both tasks, the associations are mode independent. Proctor and Vu (2010) showed that the effects of mixing were reduced considerably when each mapping or task used distinct key presses on the left and right hands. The relative lack of influence of mixing on the SRC and Simon effects when the tasks have unique responses implies that suppression of direct activation of the corresponding response occurs primarily when tasks share responses.

We have conducted experiments with members of the MURI team from Carnegie Mellon University using an expanded mixing paradigm that includes situations in which both tasks and mappings are mixed and in which payoffs and proportions of different trial types are manipulated. The purpose of the project is to model two major aspects of task performance, practice and sequential effects, by using an ACT-R modeling environment (Dutt, Gonzalez, Yamaguchi, & Proctor, 2010). For the experiments, two types of tasks could occur on any trial, an SRC task where subjects responded to the locations of visual stimuli and a Simon task where subjects responded to the color of visual stimuli while ignoring the stimulus location. Furthermore, for the SRC task, subjects were required to respond by pressing a response key whose location was compatible with the stimulus location on some trials, and by pressing a response key whose location was incompatible

on other trials. The basic findings in the mixed-task experiment were (1) responses are faster for the Simon task than for the SRC task; (2) the practice effect is larger for the SRC task than for the Simon task; (3) overall, the SRC and Simon effects are eliminated (more specifically, they are eliminated when the spatial correspondence on the current trial is different from that on the preceding trial, but they are present when the spatial correspondence on the current trial is the same as that on the preceding trial); (4) the cost of switching the compatibility relationship is larger for the SRC task than for the Simon task; (5) the cost of switching task is larger for the Simon task than for the SRC task.

The first outcome can be attributed to the number of processing steps required for the SRC task being greater than that for the Simon task (see Figure 3): For the Simon task, subjects must identify the stimulus color and select a correct response, whereas for the SRC task, they have to identify the stimulus location, determine an appropriate S-R mapping rule, and then select a correct response. The second outcome can be attributed to improvement of the mapping determination process. The third outcome is consistent with our previous studies (Yamaguchi & Proctor, 2006). The fourth outcome is due to the fact that the compatibility relation is task-relevant for the SRC task and task-irrelevant for the Simon task, so that the influence of switching that relation is more strongly manifested for the former than the latter task; thus, the effect is due mainly to the mapping-determination stage. The last outcome is consistent with the fact that the cost of switching task is typically larger from a difficult task to an easy task than in the reverse direction. As the SRC task is more complex than the Simon task, a larger cost of task-switching is expected for the Simon task than for the SRC task.

Given these basic findings, we conducted two additional experiments where we manipulated (a) payoffs given to correct responses for the compatible- and incompatible-mapping tasks (Experiment 2) and (b) frequencies of the SRC and Simon trials (Experiment 3). In Experiment 2, half the subjects received a higher payoff for the compatible-mapping task (*C-favor group*), and the other half a higher payoff for the incompatible-mapping task (*I-favor group*). The experiment replicated (a) faster responses for the Simon task than for the SRC task and (b) the larger practice effect for the SRC task than for the Simon task. There was a dissociation between the Simon and SRC effects; the Simon effect was positive (16 ms for RT data, 1.77% for percentage error data), whereas the SRC effect was negative (-14 ms, -0.76%). Moreover, the error data suggest that in the first trial block, the compatibility effect (average of the SRC and Simon effects) was positive for the C-favor group and negative for the I-favor group, but for both groups, the effects gradually approached zero over trials. Thus, the payoff manipulation was effective at early stages, but its influence decreased and subjects seem to have performed the mixed-task in later trials just as the subjects in Experiment 1 did.

In Experiment 3, we manipulated the frequencies of occurrence of the SRC and Simon tasks: For half the subjects, 80% of trials were from the Simon task (*mostly-Simon group*), and for the other half, 80% of trials were from the SRC task (*mostly-SRC group*). For the mostly-Simon group, responses were generally faster for the Simon task than for the SRC task, but for the mostly-SRC group, responses were initially faster for the Simon task and then for the SRC task in later trials. Thus, as subjects experienced the SRC task more often than the Simon task, they became more proficient at performing the SRC task than the Simon task. In contrast to the prior experiments, the mostly-SRC group showed similar costs of switching tasks, implying that, in this case, the SRC task was no longer

more difficult than the Simon task.

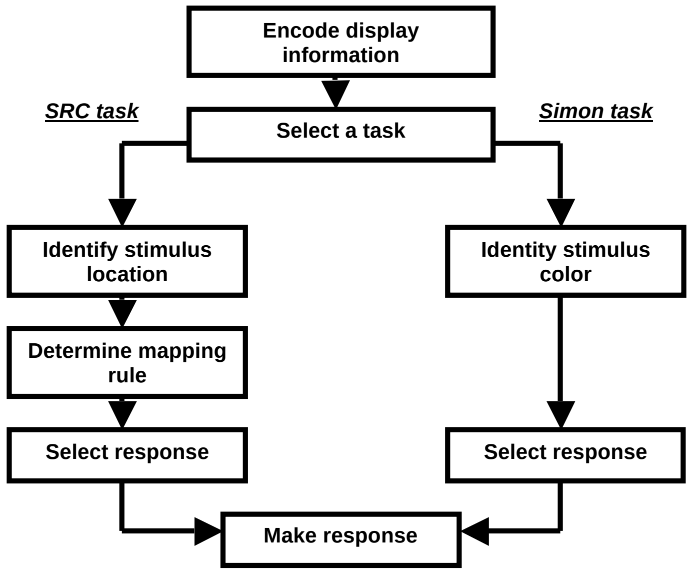


Figure 3. *A hypothetical process architecture for the expanded mixing paradigm.*

An ACT-R model of the mixed-task condition was constructed based on the Instance-Based Learning Theory (IBLT; Gonzalez, Lerch, & Lebiere, 2003) and calibrated to the data of Experiment 1 (Dutt et al., 2010). The same model was applied to the data of Experiments 2 and 3, and the results suggested a good fit of the original model without major alternations. Thus, the project suggests the usefulness of the ACT-R/IBLT model for explaining human performance under multi-tasking conditions.

4. Performance of concurrent tasks. Often, not only does one have to be prepared to perform one of two or more tasks, but multitasking demands require that the tasks be performed concurrently. The research conducted for this component examined issues relating to whether skills are acquired when attention is directed toward another task and coordination of performance across different tasks.

An issue of importance is the extent to which attention is required during learning of a skill and for that newly learned information to be expressed subsequently. We investigated this issue with an auditory version of the practice/transfer paradigm described in section 2, in which subjects practiced making spatially incompatible responses to left and right tones based on their locations and then made the same responses based on the auditory frequencies (high or low) of the tones (Miles & Proctor, 2010). The unique aspect of the study was that some participants performed the incompatible-mapping task while concurrently tracking a ball displayed on the screen by moving the computer mouse. Because the ball tracking task was attentionally demanding, participants could pay less attention to the incompatible-mapping task. Consequently, if attention is required for establishing the new S-R associations, a smaller transfer effect to the Simon task should be obtained, as compared to those participants who performed the incompatible-mapping task without the ball tracking. This is the outcome that was obtained. As in

previous research, practice with the spatially incompatible mapping eliminated the Simon effect in the transfer task when there was no concurrent task during the acquisition phase. However, the Simon effect was not reduced in the transfer session when the tracking task was performed concurrently during practice. In addition, we examined the influence of the concurrent ball tracking task in the transfer task. That is, all participants performed the incompatible-mapping task without the ball tracking and then transferred to the Simon task either with the ball-tracking or without it. The Simon effect was equivalent for the two groups, suggesting that attention is not required to manifest the effect of incompatible S-R links. These results imply that the transfer effect reflects "automatic retrieval" of the learned skills, which is consistent with the instance-based learning theory (Logan, 1988; Gonzalez et al., 2003; see Gonzalez's report).

Dual-task performance is often studied in what is called the *psychological refractory period* (PRP) paradigm, in which stimuli for two different tasks are presented in close temporal proximity, each of which requires a speeded response (see Lien & Proctor, 2002, for a review). This paradigm, which has a long history of research in applied experimental psychology much like that of compatibility effects (Telford, 1931), is of value because it allows assessment of both general attentional demands of response selection and more specific interactions across tasks. The most widely established finding is that the response for Task 2 is slowed considerably when the time between stimulus onsets is short, and this PRP effect is typically attributed to a response-selection bottleneck. One issue has been whether this bottleneck is bypassed, and the dual-task interference eliminated, when the stimuli and responses have a high form of compatibility called *ideomotor compatibility*. An example of an ideomotor compatible task is responding to spoken letter stimuli by saying each letter's name. The basic idea is that the high S-R compatibility of such tasks may allow the response to be generated automatically, without requiring the typical response-selection process.

During the training MURI, we conducted two studies examining the PRP effect with ideomotor compatible tasks. Shin, Cho, Lien, and Proctor (2007) reported three experiments in which both Task 1 and Task 2 were two-choice tasks: Task 1 required manual responses (keypresses or joystick movements) to left and right pointing arrows presented in left and right locations, respectively, and Task 2 required vocal naming responses to letters. Shin and Proctor (2008) varied whether the first task had two or four choices, also in three experiments. A PRP effect for Task 2 response time was evident in all of the conditions of these two studies, showing that ideomotor tasks do not seem to bypass the response-selection bottleneck. Of most concern for present purposes is that across four or more dual-task blocks of up to 48 trials each in all experiments, only in one case, that of auditory-vocal Task 1 and visual-joystick Task 2 (Shin & Proctor, 2008), did the PRP effect decrease with practice, and even there it was still evident in the last trial block. In fact, for the two experiments in which Task 1 used joystick responses to visual stimuli (Shin et al., Experiment 2; Shin & Proctor, Experiment 1), the PRP effect increased across blocks. So, even with very highly compatible individual tasks, practice is not sufficient to overcome dual-task interference.

We also conducted studies that used the PRP paradigm to examine cross-talk between spatial tasks performed with the left and right hands. For these experiments, the stimulus locations for Task 1 were to the left of center and those for Task 2 were to the right of center, and the responses were made with fingers on the left and right hands

respectively. Each task had the same number of alternatives, two in the experiments of Vu and Proctor (2006), three in the experiments of Proctor and Vu (2009b), and four in those of Proctor and Vu (2009a). The main variable of interest in all cases was the consistency of mappings for the two tasks. Mappings were consistent when both were compatible or both incompatible (e.g., make the mirror opposite response) and inconsistent when Task 1 used one mapping and Task 2 another. In all cases, a benefit for consistent mappings was obtained, similar to that reported initially by Duncan (1979) for three-choice tasks. However, the basis for this consistency benefit was different for the two-choice tasks when compared to those involving more than two choices. For two-choice tasks, several findings (e.g., presence of benefit mainly at short onset intervals; no benefit when one task used auditory stimuli) implied that the consistency benefit was due to an emergent perceptual blank feature that allowed subjects to respond compatibly to blank regions of the visual display (i.e., when both task mappings were incompatible, the responses for both tasks corresponded to the locations in which stimuli did not occur). For 3- and 4-choice tasks, in contrast, the evidence favored Duncan's original hypothesis that the benefit comes about from having only a single mapping rule to apply to both tasks, rather than having to choose between rules. These results suggest that performance will be best when consistency of mappings is maintained across tasks and that training that highlights consistent relationships may be most beneficial.

Finally, a characteristic of multitasking in many situations is that a person must determine how much effort to devote to a particular task and when to switch attention from one task to another. We examined issues relating to this strategic aspect of multitasking in a synthetic work environment (Wang, Proctor, & Pick, 2007, 2009) intended to be a generic representation of a variety of multitasking situations. This environment requires concurrent performance of four tasks (math, memory search, visual monitoring, and auditory monitoring), each represented in a quadrant of the computer screen, that require positioning of a cursor with a computer mouse on a response button, and then clicking on the button. Points are received for correct responses and lost for incorrect responses, and the goal is to maximize the number of points obtained. We varied the payoffs for the two more cognitively demanding tasks, math and memory search, jointly (Wang et al., 2007) or singly (Wang et al., 2009) between participants to determine sensitivity of strategies to the payoff schedule across eight 5-min sessions. Participants were sensitive to the payoff differences, performing a task relatively more when its payoff was high than when it was low. When the payoffs for the math and memory task were varied concurrently, performance of both tasks reflected their relative emphasis. However, when the payoff was varied explicitly for only one of the tasks, implicitly modifying the relative payoff for the other, just performance of the task associated with the explicit payoff was affected. For the next four transfer sessions, the payoff schedule was switched for half of the participants and kept the same for the other half. Results showed that the participants modified their strategies consistent with the new payoffs. However, residual effects of prior payoffs were evident such that the performance of the subjects for whom the payoff schedule changed did not match that of subjects who had performed with that payoff schedule all along. General implications of this research include that payoffs for multiple-task environments need to be explicit, and practice should be provided for strategy development. When payoffs change, strategies adopted reflect current and previous payoffs.

5. Summary. Our research has shown that there are benefits of applying individual principles in the training of specific tasks. However, this training is not isolated and can suffer from interference from components within a task or between tasks. We have identified specific factors that influence the learning and transfer of S-R associations and how they are impacted by task switching and multitasking.

References

Alluisi, E. A., & Warm, J. S. (1990). Things that go together: A review of stimulus-response compatibility and related effects. In R. W. Proctor & T. G. Reeve (Eds.), *Stimulus-response compatibility: An integrated perspective* (pp. 3-30). Amsterdam: North-Holland.

Cohen, A. L., Bayer, U. C., Jaudas, A., & Gollwitzer, P. M. (2008). Self-regulatory strategy and executive control: Implementation intentions modulate task switching and Simon task performance. *Psychological Research*, 72, 12-26.

Duncan, J. (1979). Divided attention: The whole is more than the sum of its parts. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 216-228.

Dutt, V., Gonzalez, C., Yamaguchi, M., & Proctor, R. W. (2010). *Instance-based learning models of SRC and Simon effects*. Manuscript submitted for publication.

Dutta, A., & Proctor, R. W. (1992). Persistence of stimulus-response compatibility effects with extended practice. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 801-809.

Fitts, P. M., & Deininger, R. L. (1954). S-R compatibility: Correspondence among paired elements within stimulus and response codes. *Journal of Experimental Psychology, 48*, 483-492.

Fitts, P. M., & Seeger, C. M. (1953). S-R compatibility: Spatial characteristics of stimulus and response codes. *Journal of Experimental Psychology, 46,* 199-210.

Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science, 27*, 591-635.

Lien, M.-C., & Proctor, R. W. (2002). Stimulus-response compatibility and psychological refractory period effects: Implications for response selection. *Psychonomic Bulletin & Review, 9*, 212-238.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95*, 492-527.

Miles, J. D., & Proctor, R. W. (2008). Improving performance through implementation intentions: Are preexisting response biases replaced? *Psychonomic Bulletin & Review, 15,* 1105–1110.

Miles, J. D., & Proctor, R. W. (2010). *Attention is required for the acquisition but not expression of new spatial associations*. Manuscript submitted for publication.

Proctor, R. W., & Lu, C.-H. (1999). Processing irrelevant location information: Practice and transfer effects in choice-reaction tasks. *Memory & Cognition, 27,* 63-77.

Proctor, R. W., & Vu, K.-P. L. (2002). Mixing location irrelevant and relevant trials: Influence of stimulus mode on spatial compatibility effects. *Memory & Cognition, 30*, 281-294.

Proctor, R. W., & Vu, K.-P. L. (2009a). Determinants of the benefit for consistent stimulus-response mappings in dual-task performance of four-choice tasks. *Attention, Perception & Psychophysics, 71*, 734-756.

Proctor, R. W., & Vu, K.-P. L. (2009b). Determinants of the benefit for consistent stimulus-response mappings in dual-task performance of three-choice tasks. *Attention, Perception, & Psychophysics, 71*, 1771-1781.

Proctor, R. W., & Vu, K.-P. L. (2009c). Task-defined associations are mode specific for selection of relevant dimension but mode independent for selection of mapping. *Quarterly Journal of Experimental Psychology, 62*, 371-392.

Proctor, R. W., & Vu, K.-P. L. (2010). Stimulus-response compatibility for mixed mappings and tasks with unique responses. *Quarterly Journal of Experimental Psychology, 63,* 320–340.

Proctor, R. W., Vu, K.-P. L., & Marble, J. G. (2003). Eliminating spatial compatibility effects for location-relevant trials by intermixing location-irrelevant trials. *Visual Cognition, 10,* 15-50.

Proctor, R. W., & Wang, H. (1997). Differentiating types of set-level compatibility. In B. Hommel & W. Prinz (Eds.), *Theoretical issues in stimulus-response compatibility* (pp. 11-37). Amsterdam: North-Holland.

Proctor, R. W., Yamaguchi, M., & Vu, K.-P. L. (2007). Transfer of noncorresponding spatial associations to the auditory Simon task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 245-253.

Proctor, R. W., Yamaguchi, M., Zhang, Y., & Vu, K.-P. L. (2009). Influence of visual stimulus mode on transfer of acquired spatial associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 434-445.

Shin, Y. K., Cho, Y. S., Lien, M.-C., & Proctor, R. W. (2007). Is the psychological refractory period effect for ideomotor compatible tasks eliminated by speed-stress instructions? *Psychological Research/Psychologische Forschung, 71,* 553-567.

Shin, Y. K., & Proctor, R. W. (2008). Are spatial responses to visuospatial stimuli and spoken responses to auditory letters ideomotor-compatible tasks? Examination of set-size effects on dual-task interference. *Acta Psychologica, 129,* 352-364.

Simon, J. R. (1990). The effects of an irrelevant directional cue on human information processing. In R. W. Proctor & T. G. Reeve (Eds.), *Stimulus-response compatibility: An integrated perspective* (pp. 31-86). Amsterdam: North-Holland.

Tagliabue, M., Zorzi, M., Umiltà, C., & Bassignani, F. (2000). The role of LTM links and STM links in the Simon effect. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 648-670.

Telford, C. W. (1931). Refractory phase of voluntary and associative response. *Journal of Experimental Psychology, 14,* 1-35.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*, 327-352.

Vu, K.-P. L. (2007). Influences on the Simon effect of prior practice with spatially incompatible mappings: Transfer within and between horizontal and vertical dimensions. *Memory & Cognition, 35,* 1463-1471.

Vu, K.-P. L, & Proctor, R. W. (2004). Mixing compatible and incompatible mappings: Elimination, reduction, and enhancement of spatial compatibility effects. *Quarterly Journal of Experimental Psychology, 57A*, 539-556.

Vu, K.-P. L., & Proctor, R. W. (2006). Emergent perceptual features in the benefit of consistent stimulus-response mappings on dual-task performance. *Psychological Research, 70,* 468-483.

Vu, K.-P. L, Proctor, R. W., & Urcuioli, P. (2003). Transfer effects of incompatible location-relevant mappings on subsequent visual or auditory Simon tasks. *Memory & Cognition, 31*, 1146-1152.

Wang, D.-Y. D., Proctor, R. W., & Pick, D. F. (2007). Acquisition and transfer of attention-allocation strategies in a multiple-task work environment. *Human Factors*, *49*. 995-1004.
Wang, D. D., Proctor, R. W., & Pick, D. F. (2009). Allocation of effort as a function of payoffs for individual tasks in a multitasking environment. *Behavior Research Methods, 41*, 705-716.
Yamaguchi, M., & Proctor, R. W. (2006). Stimulus-response compatibility with pure and mixed mappings in a flight task environment. *Journal of Experimental Psychology: Applied, 12*, 207-222.
Yamaguchi, M., & Proctor, R. W. (2009). Transfer of learning in choice reactions: Contributions of specific and general components of manual responses. *Acta Psychologica, 130,* 1–10.

The MURI training taxonomy

William D. Raymond, Alice F. Healy, and Lyle E. Bourne, Jr.

University of Colorado, Boulder

The MURI training taxonomy

William D. Raymond, Alice F. Healy, and Lyle E Bourne, Jr.

University of Colorado, Boulder

The goal of the Training MURI is to quantify effects on performance of different training methods for complex military tasks. However, the range of variables that can affect training and the multiplicity of tasks that may require training prevent an exhaustive quantification of training effects for specific tasks and training scenarios. To render the study of training effects tractable and to guide research, both in this MURI and in future work, we have developed a taxonomy that includes separate dimensions for task description, training procedure, and the context and assessment of task performance. The taxonomy, described in this paper, provides a framework by which training effects can be assessed and predicted componentially for any task. Examples of its application are discussed for specific laboratory tasks.

1. Introduction. The goal of the Training MURI is to quantify the effects on performance of different training methods for complex military tasks. Our multi-pronged approach in meeting this goal has involved extensive basic experimental research exploring the effects of training variables on performance in laboratory tasks, together with computational modeling of human task performance. The empirical research is the basis for a set of training principles that relate training methods and outcomes and can assist in the development of training regimens by the military. However, the range of variables that can affect training efficacy and the multiplicity of tasks that may require training prevent an exhaustive quantification of training outcomes for specific tasks and training scenarios. In order to render the study of training effects tractable and to guide research, both in this MURI and in future work, we have developed a multi-dimensional taxonomy, which will provide a framework by which training effects can be assessed and predicted for any task.

A taxonomy is a hierarchical classification based on a consistent set of principles that can be tested for agreement with empirical data and whose order corresponds to a real order of the classified elements (Krathwohl, Bloom, & Masia, 1964). To be testable, features of the MURI taxonomy should thus be relatable to the design of laboratory experiments being conducted to explore training variables in the MURI. That is, the taxa of the three dimensions must be capable of capturing the tasks, manipulations, and measured responses of the experiments. At the same time, taxa should be no finer than the experimental manipulations. In addition, the features should be broad enough to cover task, training, and performance requirements that may likely be encountered in a military context, which may be broader than the scope of current experimental coverage (although military tasks frequently include the experimental tasks as subtasks). Of further interest to the military is relating taxon effects captured by the MURI taxonomy to the task

taxonomy in the military's Improved Performance Integration Tool (IMPRINT; Archer et al., 1999) simulation software. Thus, a further constraint on the taxonomy is that there be a mapping from MURI task taxa to IMPRINT task taxa.

At the highest level, as specified by the MURI grant proposal, the taxonomy we have developed involves a four-dimensional decomposition of the training space. It includes separate dimensions of classification for task description, training procedure, and the context and assessment of task performance. The training principles are considered the fourth dimension. The first three dimensions are structured as hierarchical feature decompositions whose values and relationships are described in this paper.

An assumption of the decompositional approach is that the goal of predicting performance for any task can be accomplished by combining the effects on each performance measure of individual training components for all task elements. Accomplishing this goal depends on an exploration of the matrix of cells in the training space defined by the taxa of the three dimensions. This work extends beyond the MURI; however, the space has been partially explored by empirical studies we have conducted, and identification of current coverage allows for planning of future work.

This report presents a brief review of approaches to taxonomies in each of the three dimensions, together with motivation and description of the taxa selected for use in the MURI taxonomy. Principles used to select taxa, as well as the correspondence between the organization of taxa and the phenomena they are meant to capture, are highlighted. After presenting the taxonomy, application of it to two tasks, a digit data entry task (see, e.g., Healy, Kole, Wohldmann, Buck-Gengler, & Bourne, in press) and a visual search task (Young, Healy, Gonzalez, Dutt, & Bourne, in press), is discussed to illustrate how a taxonomic analysis can facilitate our understanding of task acquisition. A taxonomic analysis using IMPRINT task taxa and MURI training and performance taxa has been performed on all experimental tasks conducted in conjunction with the MURI. The analyses have been compiled to produce a planning matrix that shows the current extent to which the training space has been investigated and that can be used to plan future research. Finally, areas that we have identified as needing further development to enhance taxonomic analysis of the training space are discussed.

2. Task type. A general definition of a task was given by Miller (1953) to accommodate the analysis of increasingly complex human activities. According to Miller, a task is "a group of discriminations, decisions and effector activities related to each other by temporal proximity, immediate purpose and a common man-machine output" (cited in Meister, 1976, p. 96). The definition can be interpreted as recognizing that tasks involve perceptual inputs, cognitive processing, and motor responses. From this starting point, the development of a specific taxonomy of human tasks has been approached in a variety of ways, including classifications based on task stimuli, human behavior during task performance, or human ability requirements (see Companion & Corso, 1982). The approach to classification clearly depends on the purpose to which a taxonomy is to be put (see Gawron, Drury, Czaja, & Wilkins, 1989).

One class of task taxonomies particularly important in the fields of human learning and performance begins with the notion that tasks can be analyzed according to their demand on human abilities (see Fleishman, 1978). Roth (1992) proposed a taxonomy with five broad ability taxa: attentional, perceptional, psychomotor, physical,

and cognitive. As an application of the taxonomy, empirical data are used by Roth (1992) to relate the effects of external stressors to each ability taxon. Weighted decompositions of specific subtasks are then available to predict stressor effects at the task level.

The task decomposition adopted for the MURI, shown in Table 1, builds on taxonomies like the Roth (1992) taxonomy of abilities, introducing a finer classification of abilities, while keeping the number of taxa tractable. Taxa are selected principally to capture the cognitive processing of stimuli. Categorizing information processing tasks was considered to be central, because of both the military's primary desire to optimize training for the networked battlefield and the fact that most empirical studies conducted for the MURI have largely been designed to explore cognitive processing, with concomitant perceptual and psychomotor processes. In information processing tasks inputs are initially processed using perceptual and attentional abilities. Information is further synthesized with higher-order cognitive processes and memory, and output responding is planned. Finally, a psychomotor response in produced. This sequential processing cycle is reflected in the hierarchy of the taxonomy.

Table 1. *The MURI task dimension*.

<table>
<tr><td rowspan="7" colspan="2">Perceptual/Attentional Processing</td><td>Visual detection</td></tr>
<tr><td>Visual discrimination</td></tr>
<tr><td>Language processing (written)</td></tr>
<tr><td>Auditory detection</td></tr>
<tr><td>Auditory discrimination</td></tr>
<tr><td>Language processing (oral)</td></tr>
<tr><td>Haptic processing</td></tr>
<tr><td rowspan="9">Cognitive/Affective Processing</td><td rowspan="7">Synthesis</td><td>Executive control/Monitoring</td></tr>
<tr><td>Memory/Symbolic representation</td></tr>
<tr><td>Imagery/Visual representation</td></tr>
<tr><td>Concept formation/Classification</td></tr>
<tr><td>Reasoning/Problem solving</td></tr>
<tr><td>Decision making</td></tr>
<tr><td>Motivation/Affect</td></tr>
<tr><td rowspan="2">Response Planning</td><td>Language planning</td></tr>
<tr><td>Motor response planning</td></tr>
<tr><td rowspan="3" colspan="2">Physical/Communicative Response</td><td>Manipulation/Fine motor output</td></tr>
<tr><td>Action/Gross motor output</td></tr>
<tr><td>Language production</td></tr>
</table>

Although the current task taxonomy is sufficiently comprehensive to decompose MURI laboratory tasks, it may be that the use of the task taxonomy for some Army tasks may require additional distinctions. New ability taxa could readily be incorporated into the existing taxonomy. In addition, it may be desirable to allow for the inclusion of the relative contribution of each taxon to the performance of a task, which may vary from task to task and also across training.

The MURI task taxa are different from the task taxa used for military simulation in IMPRINT; however, it is possible to establish a mapping between the MURI features

and the IMPRINT task taxa, although the mapping is not one-to-one. The mapping is shown in Table 2.

3. Training method. The training dimension covers variables that capture the method of instruction and the types of activities performed during learning. Berliner (1983) recognized the need for more rigorous definitions of educational treatments, and he provided a taxonomy for classroom activity structures that takes into account the roles of students and teachers in instruction, classroom group size, response and feedback types, and the range and source of content.

Table 2. *Mapping between MURI task taxa and IMPRIMT task taxa*.

MURI task taxa	**IMPRINT task taxa**
Visual detection, Visual discrimination	Visual
Language processing (written)	Communication (reading & writing)
Auditory detection, discrimination	*(no corresponding IMPRINT taxon)*
Language processing (oral)	Communication (oral)
Haptic processing	Fine motor - discrete
	Fine motor - continuous
Executive control/Monitoring	Information processing
Memory/Symbolic representation	Information processing
	Communication (oral)
	Communication (reading & writing)
Imagery/Visual representation	Information processing
Concept formation/Classification	Information processing
Reasoning/Problem solving	Information processing
	Numerical Analysis
Decision making	Information processing
Motivation/Affect	*(no corresponding IMPRINT taxon)*
Language planning	Communication (oral)
Motor response planning	Communication (reading & writing)
	Fine motor - discrete
	Fine motor - continuous
Manipulation/Fine motor output	Fine motor - discrete
	Fine motor - continuous
Action/Gross motor output	Gross motor - light
	Gross motor - heavy
Language production	Communication (reading & writing)
	Communication (oral)

A broader perspective of training methods is captured by Jonassen and Tessmer (1996/97), who present a taxonomy of instructional and learning strategies and specific tactics for achieving training outcomes. Their strategies, compiled from a review of relevant literature, range from traditional objective strategies (e.g., present examples, provide practice, provide feedback) to more outcome-specific approaches (e.g., model

cognitive activity, relate to prior knowledge, scaffold performance). Although the strategy set presented may cover a large proportion of training scenarios, additional detail is desirable for decomposing training scenarios.

There are two major pieces in the decomposition of task learning in the MURI taxonomy: pedagogy and practice. Pedagogy captures the method of task instruction. The pedagogy taxa are shown in Table 3, along with the values each parameter may assume. The practice taxa are used to describe the nature of practice performed during training. Practice can be further subdivided into scheduling parameters, task parameters, feedback parameters, and training context parameters. The parameter groupings for the practice taxa and the currently defined parameters within each grouping are shown in Table 4. Standard parameter values are indicated as default values in Tables 3 and 4, with the range of alternative values indicated.

Table 3. *The MURI training dimension pedagogy taxa.*

Pedagogy parameters	Instruction method	Lecture/Instruction
		Demonstration
		Discovery
		Computer instruction
		Simulation (i.e., interaction with computerized representation of a task)
		Modeling (mimicking = observe and mimic a model performing the task)
	Discussion/Question & answer	default = 1-way; 2-way
	Immersion	default = no; yes (embedded in field context)
	Learning location	default = local; remote or “distance learning”
	Individualization	default = no; yes - e.g., human or intelligent computer tutoring
	Group training	default = no; group size.
	Automation	default = no; yes.

Evidence for effects of parameters from both groupings on skill acquisition in a variety of tasks has been demonstrated in numerous laboratory studies (see Proctor & Vu, 2006 for a review; see also O’Neil, 2003, on distance learning; Carpenter, Pashler, Wixted, & Vul, 2008, and Szpunar, McDermott, & Roediger, 2008, on testing during training). Corroborative evidence comes from studies of expert performance. Although the set of parameter values selected for inclusion in the MURI taxonomy are intended to allow an analysis of most training scenarios, additional pedagogy and practice parameters may be added to the taxonomy when they become necessary.

4. Performance context and assessment. Taxonomies of training criteria have been important in assessing the effectiveness of training programs in the business environment. A simple and influential taxonomy of assessment criteria (Kirkpatrick, 1987; see Alliger, Tannenbaum, Bennett, Traver, & Shotland, 1997, for an augmented version of the

taxonomy) specifies four categories of criteria: reactions, learning, behavior, and results. The category of reactions assesses a trainee's judgment of training usefulness, difficulty, and pleasantness. Learning encompasses all post-test assessments of knowledge and skill, although tests most commonly measure declarative knowledge of training materials. The behavior category captures on-the-job performance or behavior. The results category includes measures of the organizational impact of training.

Table 4. *The MURI training dimension practice taxa*.

Practice parameters	Scheduling Parameters	Number of items/trials	
		Item difficulty	default = unspecified; difficulty level
		Item repetition	default = massed; repetition interval
		Time spacing	default = no rest; rest interval
		Distribution	default = mixed; blocked
		Change in spacing	default = none; expansion; contraction
		Session (parameters of importance; at least number of sessions and session spacing)	
		Testing	default = no testing; test schedule
		Overlearning	default = no; yes
	Task Parameters	Scope	part, e.g., mental rehearsal; default = whole; supplemental
		Deep processing	default = no; yes
		Mediation (e.g., use of prior knowledge)	default = no; yes
		Attentional focus	default = no focus; internal, external
		Attentional breadth	default = intermediate; global, local
		Stimulus-response compatibility	default = yes; no
		Mapping type	default = consistent; varied
		Contralateral training	default = no; yes
		Time pressure	default = no; yes
		Stressor	default = no; yes
	Feedback Parameters	Presence of (response) feedback	default = no; yes
		Feedback scheduling (relative to items)	
	Training Context Parameters	Distractor	default = no; yes
		Secondary activity	default = none; simultaneous; sequential

Of importance to the current research effort from this taxonomy are the categories of behavior and learning, that is, measures of performance on the job (i.e., "in the field") and of post-test performance. However, the Kirkpatrick (1987) taxonomy lacks sufficient detail to apply it to specific training situations. The behavior category does not capture differences between training and performance environments, which are known to impact performance. Additionally, the learning category in the Kirkpatrick taxonomy leaves

unspecified what types of measures may be necessary to assess training outcomes. The performance dimension of the MURI taxonomy incorporates these two components with separate taxa, of performance context and of performance assessment, but provides greater detail. Performance context covers the conditions of and delay to post-training performance, relative to training; performance assessment specifies measures of performance.

4.1. Performance context. The performance context component relates the environment of post-training performance to the training environment. The major component of performance context captures the relationship of performance to the items, context, and task encountered in training. In addition, performance context is concerned with the time since training and the frequency of any intervening refresher training prior to performance. The taxa in the MURI taxonomy for performance context are shown in Table 5.

Table 5. *Decomposition of the performance context dimension of the MURI taxonomy.*

<table>
<tr><td rowspan="3">Transfer parameters</td><td>New items, item order, or item distribution</td><td>default = same as training; different items, order, or distribution</td></tr>
<tr><td>New context</td><td>default = same as training; different context</td></tr>
<tr><td>New task</td><td>default = same as training; different task</td></tr>
<tr><td>Retention interval</td><td colspan="2">default = none; time since training</td></tr>
<tr><td>Refresher training schedule</td><td colspan="2">default = none; refresher schedule</td></tr>
</table>

4.2. Performance assessment. Complex training goals can be evaluated using systems designed to facilitate assessment of the acquisition of knowledge, such as in the taxonomy of cognitive learning developed by Bloom, Englehart, Furst, Hill, and Krathwohl (1956). In their taxonomy, cognitive learning goals can be arranged in a hierarchy of knowledge complexity. Mastering any level of the hierarchy requires mastery of the behaviors in the taxa below it. The levels proposed by Bloom et al. are shown in Table 6, along with methods of assessment for each level.

Table 6. *The Bloom et al. (1956) taxonomic hierarchy for the cognitive learning domain.*

Learning Goal	Assessment
Knowledge	Recall or recognize information
Comprehension	Comprehend or interpret information
Application	Use information to complete a task
Analysis	Distinguish, classify, and relate knowledge
Synthesis	Originate and combine ideas
Evaluation	Appraise and assess ideas based on standards

The Bloom et al. (1956) taxonomy focuses on the acquisition of verbal, or declarative, knowledge and associated behaviors. Skill performance can generally be objectively assessed in terms of speed or accuracy of task completion. Separate measures are needed, because it has been shown that there are tradeoffs between speed and accuracy in some tasks. In a data entry task, speed and accuracy show different patterns of results; speed improves with training while accuracy declines (Healy, Kole, Buck-Gengler, & Bourne, 2004). However, in other scenarios, the opposite pattern might obtain. Moreover, situations in which training produces improved efficiency of performance (i.e., faster and more accurate responding) need to be differentiated from those in which it alters only the speed-accuracy criterion. It is also important to assess performance on sub-components of a task. For example, the response times for executing the different steps of a digit data entry task are not always positively correlated, with typers slowing down on one step in order to be faster on another (Healy et al., 2004).

In some tasks, there is also a necessity to develop some index of changes in the learner's cognition during training. For example, in a binary classification task, Bourne, Raymond, and Healy (2010) have shown that even when both speed and accuracy measures show continuous improvement, subjects use different strategies to guide their responses, often changing strategies during training. Measures must be developed to assess changes in cognitive strategies, because the strategy chosen may impact speed and accuracy, or even retention and transfer.

Table 7. *The Kraiger, Ford, and Salas (1993) classification of learning outcomes and associated measures of assessment.*

<table>
<tr><th colspan="3">Learning outcome</th><th>Assessment</th></tr>
<tr><td rowspan="3">Cognitive Outcomes</td><td colspan="2">Verbal Knowledge</td><td>Tests of memory</td></tr>
<tr><td colspan="2">Knowledge Organization</td><td>Probe cognitive structures</td></tr>
<tr><td colspan="2">Cognitive Strategies</td><td>Probe task protocol</td></tr>
<tr><td rowspan="3">Skill-based Outcomes</td><td rowspan="2">Compilation</td><td>Proceduralization</td><td rowspan="2">Change in performance</td></tr>
<tr><td>Composition</td></tr>
<tr><td colspan="2">Automaticity</td><td>Test with interference stimuli or distractors</td></tr>
<tr><td rowspan="2">Affective Outcomes</td><td colspan="2">Attitudinal</td><td>Self-report</td></tr>
<tr><td>Motivational</td><td>Disposition</td><td>Self-report with increasing problem difficulty</td></tr>
</table>

Researchers have also expanded the scope of learning outcomes to include affective or attitudinal learning goals as well as knowledge and skill acquisition. Drawing on all three areas of research, Kraiger, Ford, and Salas (1993) proposed a more comprehensive taxonomy of learning outcomes, shown in Table 7. They define learning as changes in cognitive, skill-based, and attitudinal states and discuss how learning in each category can be measured (see Table 6). The Kraiger et al. (1993) classification forms the basis for the MURI performance assessment taxonomy. However, speed and accuracy measures of individual components can be combined with the different levels to form a taxonomy of assessment tests. Having quantified the outcome of a particular

training scenario, the effectiveness of training can be measured by comparing post-training performance with performance before or at the beginning of training, using an accepted measure of training, such as the training effectiveness ratio (Wickens & Holland, 2000). Performance results can then feed back to further training design.

5. Using the taxonomy. A taxonomic breakdown of task, training, and performance dimensions provides a way to explore the training space incrementally. By holding the task constant, training effects can be quantified within many cells in the taxonomic space across the training and performance dimensions. Empirical data are generated by experimentation, with various separate experimental manipulations providing speed, accuracy, and strategy measures of performance for the effects of many training and performance contexts on task taxa. As examples of this approach, we will consider the coverage provided by experiments using two tasks, a simple number typing task (digit data entry) and a more complex visual search task (the RADAR task).

Digit data entry is one simple task that has been extensively used by the MURI investigators to explore the effects of training on skill acquisition (e.g., Healy et al., in press). Most basically, the digit data entry task consists of typing, using the number keypad, a series of four-digit numbers presented visually on a computer screen. In this form, the task can be broken down, using the task taxonomy, into four MURI taxa: Visual detection (reading numbers from the screen), Memory/Symbolic representation (the cognitive representation of each number), Motor response planning (for typing each number), and Manipulation/Fine motor output (typing).

Pedagogy in all digit data entry experiments simply involved (written) instruction. Practice in all training scenarios involved the repeated entry of numbers. However, experiments have explored the effects of varying practice scheduling parameters, including the number of items, item difficulty (e.g., by varying numerical structure or by requiring generation of numbers to be entered arithmetically), item repetition, item distribution, and the number of training sessions. Various task parameters have also been manipulated, including task scope (full typing task vs. mental rehearsal), processing depth (numeral vs. verbal presentation format), processing mediation (association of numbers with prior knowledge), contralateral training, and the presence of a physical stressor during training (hand weights). Additionally, the presence of feedback has been manipulated, as well as use of a simultaneous secondary task (articulatory suppression) and a sequential secondary task (calculation of the typing termination key). Finally, performance context has been varied from training context in terms of transfer parameters (new vs. old numbers, mental vs. physical typing task, typing hand, and typing on keypad vs. number row), post-training retention interval, and refresher training schedule.

A number of important findings are the result of analyzing task performance in terms of its component taxa for digit data entry. Measuring speed and accuracy separately revealed that these measures show different patterns of results, as noted. Moreover, different training methods can influence the results of the measures independently, with, for example, the presence of a secondary task requirement (the calculation of the typing termination key) providing a cognitive antidote to the otherwise observed decline in typing accuracy across practice (Kole, Healy, & Bourne, 2008). The scope of practice (whole task vs. mental rehearsal) has an effect on the transfer of performance, with mental practice improving retention and transfer by strengthening an effector-

independent representation (Wohldmann, Healy, & Bourne, 2008). A taxonomic analysis of the digit data entry task has also allowed us to quantify differential effects of training on individual taxa. In particular, repeated practice results in faster performance; however, the rate of improvement differs for the cognitive and motoric components of the task, with more learning occurring for the cognitive component (Healy et al., 2004).

The RADAR task, developed by Gonzalez and Thomas (2008), is a visual search task in which subjects look for symbol targets in four squares moving from the four corners to the center of a radar-like display in a fixed amount of time. Each search opportunity is called a frame. Different sets of target and distractor symbols may be shown in the squares in each of seven frames comprising a trial, and the target symbols may differ from trial to trial. The size of the target memory set includes either one or four symbols. Squares may also be blank, and there is at most one target shown per trial. Subjects are to respond only if a target in the current memory set appears in one of the squares, and scoring is on both accuracy and correct response speed. The task can be broken down into six MURI taxa: Visual detection (scanning for symbols), Memory/Symbolic representation (remembering targets in memory set), Imagery/Visual representation (of symbols seen in a frame), Decision making (target decision), Motor response planning, and Manipulation/Fine motor output (button push on detection).

Several experiments have explored the RADAR task (e.g., Young et al., in press). Pedagogy in all RADAR experiments involved (written) instruction. Practice involved repeated searches, with blocked practice of items varying in difficulty of mapping type (consistent vs. varied mapping) and processing load (size of the memory set). Training involved two sessions, and the presence of both a simultaneous secondary task (concurrent tone counting) and a sequential secondary task (action firing decision) was manipulated.

Analysis of RADAR experimental results showed that practice enhanced correct target detection times at delayed test. Analyzing speed and accuracy measures separately showed improvement in target detection accuracy (viz., fewer false alarms) with practice, but no improvement in target detection times. At training, both simultaneous and sequential secondary tasks increased correct response times, and the sequential secondary task also lowered accuracy (resulting in more missed targets). The effects on test performance of training with a secondary task depended on the nature of the secondary task. There was a detrimental effect on target detection accuracy at test (more missed targets) of training with the simultaneous secondary task, but a beneficial effect on target detection accuracy at test (fewer missed targets) of training with the sequential secondary task. These results corroborate the proposal that not all added task difficulty during training enhances task performance at test; only some difficulties are desirable during training (Bjork, 1994).

6. Possible expansions to the taxonomy. One important factor that is known to affect learning but that is not currently taken into account in the MURI taxonomy is individual differences in abilities and backgrounds. Whether or not practice in a skill makes individuals more similar or more different depends on the task, on individual differences in ability, and on individual differences in prior knowledge (Ackerman, 2007). For example, for tasks that depend on declarative knowledge, performance levels depend on whether the tasks are “open” or “closed.” Closed tasks are those that are bounded by a

reasonably finite domain of knowledge, whereas open tasks consist of those that increase with complexity. Thus, for open tasks (but not for closed tasks) there will be an increasing difference between the levels of the highest- and lowest-performing people. For tasks that allow individuals to build on existing knowledge, individual differences in prior knowledge have a larger effect on the acquisition of new knowledge than do individual differences in working memory (e.g., see Beier & Ackerman, 2005). Thus, understanding the effects of individual differences on training ultimately depends on the identification and effective use of a taxonomy of individual differences. As an example, work within the MURI has indicated that individual differences in general intelligence interact with automation, with reduced influence of general intelligence under higher levels of automation (Clegg & Heggestad, 2010). How individual differences affect training and interact with other training variables remains to be fully explored.

Group training is another important area for future work. Many Army tasks involve the interaction of multiple individuals, who share in the responsibility of task completion. Shute, Lajoie, and Gluck (2000) provide a discussion of a taxonomy of common group training techniques and the interaction of techniques with individual differences in ability, demographics, and background.

7. Toward improving training effectiveness. As the previous section indicates, experimental work performed as part of the MURI project has provided empirical data on a substantial number of task, training, and performance taxa combinations. Taking into account all MURI experiments increases the number of cells of the training space for which empirical data have been collected. To provide a basis for future research planning by the Army, we have compiled a matrix of training and performance taxa against the IMPRINT task taxa. The cells of the matrix for which empirical data have been collected are indicated with the name of the appropriate experimental task. This planning matrix is presented in Appendix A.

The number of cells in the taxonomic space defined by the MURI taxonomy outlined in this paper is large, and so at this time many cells in the taxonomic space lack empirical data from laboratory experiments related to the MURI that can be used to quantify the effects of training. It is also important to note that the empirical data generated for many cells come from exploration of only a single task, so that their generality remains to be examined. At this point it is not known whether the effects in cells of the taxonomic space that have been quantified are additive when task, training, or performance context taxa are combined. As noted, the effects of individual differences in skill and ability also need to be taken into account. Exploration of the taxonomic space must necessarily extend beyond the MURI project. However, the taxonomic decomposition made possible by the MURI taxonomy affords the Army an approach to evaluating training effectiveness across tasks, potentially facilitating improved training in the future.

References

Ackerman, P. L. (2007). New developments in understanding skilled performance. *Current Directions in Psychological Science 16,* 235–239.

Alliger, G. A., Tannenbaum, S. I., Bennett, W., Traver, H., & Shotland, A. (1997). A meta-analysis of the relations among training criteria. *Personnel Psychology, 50,* 341-359.

Archer, R., Walters, B., Yow, A., Carolan, T., Laughery, K.R., & Gillis, P. (1999). Training as a performance shaping factor in computer generated forces. *Proceedings of the 1999 Computer Generated Forces Conferences*, Orlando, FL.

Beier, M.E., & Ackerman, P.L. (2005). Age, ability and the role of prior knowledge on the acquisition of new domain knowledge. *Psychology and Aging, 20,* 341–355.

Berliner, D. C. (1983). Developing conceptions of classroom environments: Some light on the T in classroom studies of ATI. *Educational Psychology, 18*, 1-13.

Bjork, R, A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185-205). Cambridge, MA: MIT Press.

Bloom, B., Englehart, M., Furst, E., Hill, W., & Krathwohl, D. (1956). Taxonomy of educational objectives: The classification of educational goals. *Handbook I: Cognitive domain*. New York, Toronto: Longmans, Green.

Bourne, L. E., Jr., Raymond, W. D, & Healy, A. F. (2010). Strategy selection and use during classification skill acquisition. *Journal of Experimental Psychology: Language, Memory, and Cognition, 36*, 500-514.

Carpenter, S. K., Pashler, H., Wixted, J. T., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition, 36,* 438-448.

Clegg, B., & Heggestad, E. (2010, April). *Experiments on levels of automation, individual differences, and team performance*. Paper to be presented at the 2010 Ellis-Battig Memory Symposium: Optimizing the training of knowledge and skills: A review of accomplishments from the Multidisciplinary University Research Initiative (MURI) on training, 80th Annual Convention of the Rocky Mountain Psychological Association, Denver, CO.

Companion, M. A., & Corso, G. M. (1982). Task taxonomies: A general review and evaluation. *International Journal of Man-Machine Studies, 17,* 459-472.

Fleishman, E. A. (1978). Relating individual differences to the dimensions of human tasks. *Ergonomics, 21,* 1007-1019.

Gawron, V. J., Drury, C. G., Czaja, S. J., & Wilkins, D. M. (1989). A taxonomy of independent variables affecting human performance. *International Journal of Man-Machine Studies, 31,* 643-672.

Gonzalez, C., & Thomas, R. P. (2008). Effects of automatic detection on dynamic decision making. *Journal of Cognitive Engineering and Decision Making, 2*, 328-348.

Healy, A. F., Kole, J. A., Buck-Gengler, C. G., & Bourne, L. E., Jr. (2004). Effects of prolonged work on data entry speed & accuracy. *Journal of Experimental Psychology: Applied, 10,* 188-199.

Healy, A. F., Kole, J. A., Wohldmann, E. L., Buck-Gengler, C. J., & Bourne, L. E., Jr. (in press). Data entry: A window to principles of training. In A. S. Benjamin (Ed.),

Successful remembering and successful forgetting: A festschrift in honor of Robert A. Bjork. New York: Psychology Press.
Jonassen, D., & Tessmer, M. (1996/97). An outcomes-based taxonomy for instructional systems design, evaluation, and research. *Training Research Journal, 2*, 11-46.
Kirkpatrick, D. L. (1987). Evaluation of training. In R. L. Craig (Ed.), *Training and development handbook: A guide to human resource development* (3rd ed., pp. 301-319). New York: McGraw-Hill.
Kole, J. A., Healy, A. F., & Bourne, L. E., Jr. (2008). Cognitive complications moderate the speed-accuracy tradeoff in data entry: A cognitive antidote to inhibition. *Applied Cognitive Psychology, 22*, 917-937.
Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology, 78*, 311-328.
Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1964). *Taxonomy of educational objectives: The classification of educational goals*. White Plains, NY: Longman.
Meister, D. (1976). *Behavioral foundations of system development*. New York: John Wiley & Sons.
Miller, R. B. (1953). A method for man-machine task development. *Technical Report WADC-TR-53-137*. Wright-Patterson A.F.B., Ohio: Wright Air Development Center.
O'Neil, H. F. (2003). *What works in distance learning: Guidelines*. H. F. O'Neil (Ed.). Charlotte, NC: Information Age Publishing.
Proctor, R. W., & Vu, K.-P. L. (2006). Laboratory studies of training, skill acquisition, and retention of performance. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *Cambridge handbook of expertise and expert performance* (pp. 265-286). Cambridge, UK: Cambridge University Press.
Roth, J. T. (1992). *Reliability and validity assessment of a taxonomy for predicting relative stressor effects on human task performance.* Technical Report 5060-1 prepared under contract DNA001-90-C-0139, Boulder, CO: Micro Analysis and Design, Inc.
Shute, V, J., Lajoie, S. P., & Gluck, K. A. (2000). Individualized and group approaches to training. In S. Tobias & J. D. Fletcher (Eds.) *Training and retraining: A handbook for business, industry, government, and the military* (pp. 171-207). New York: Macmillan.
Szpunar, K. K., McDermott, K. B., & Roediger, H. L., III. (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 1392-1399.
Wickens, C. D., & Holland, J. G. (2000). Memory, learning, and training. In C. D. Wickens & J. G. Holland (Eds.), *Engineering psychology and human performance* (3rd ed., pp. 241-284). New York: Harper Collins.
Wohldmann, E. L., Healy, A. F., & Bourne, L. E., Jr. (2008). A mental practice superiority effect: Less retroactive interference and more transfer than physical practice. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 823-833.

Young, M. D., Healy, A. F., Gonzalez, C., Dutt, V., & Bourne, L. E., Jr. (in press). Effects of training with added difficulties on RADAR detection. *Applied Cognitive Psychology*.

Appendix A
The IMPRINT planning matrix

	Pedagogy											
	Instruction											
IMPRINT task taxons	**Lecture/Instruction**	**Demonstration**	**Simulation (interaction with computerized representation of task)**	**Discovery**	**Computer instruction**	**Modeling (Mimicking = observe and mimic a model performing task)**	**Immersion (embeded in actual field situation of task)**	**Learning location default = local; remote (i.e., distance learning))**	**Discussion/Q&A (default = 1-way, else 2-way)**	**Individualization (default = no, yes = e.g. intelligent tutoring)**	**Group training (default = no, group size)**	**Automation (default = no, yes)**
Visual	letter detection, data entry, navigation, target finding (clockface), fusion, color naming, handwriting symbols, dart throwing		radar, tank gunner			navigation						
Numerical Analysis	pseudo-arithmetic, fire control (lecture)		radar	fire control			fire control		fire control (socratic)			Clegg pasteurizer
Information processing	letter string classification, letter detection, data entry, fact learning, mental calculation, reconstruction of order		Clegg pasteurizer, radar	Clegg pasteurizer, letter string classification, time estimation, quantity estimation, sequence learning								Clegg pasteurizer
Fine motor - discrete	data entry, sequence learning, navigation, fusion		Proctor flight simulator, tank gunner, radar			navigation						
Fine motor - continuous	target finding (clockface with mouse reversal)		Proctor flight simulator	target finding (clockface with mouse reversal)								
Gross motor - light												
Gross motor - heavy												
Communication (reading & writing)	foreign language learning, letter detection, fire control, color naming			fire control			fire control		fire control (socratic)			
Communication (oral)	navigation					navigation						

			Practice					
	Scheduling parameters (of items and sessions)							
IMPRINT task taxons	**Number of items/trials**	**Item difficulty (default = unspecified; difficulty level)**	**Item repetition (default = fixed, massed)**	**Distribution (default = mixed, blocked)**	**Time spacing (default = no rest; rest) & intervalChange in spacing (default = none, expansion, contraction)**	**Sessions (whatever parameters are important; at least number and spacing)**	**Testing (default = no; test schedule)**	**Overlearning (default = no)**
Visual	(most experiments)	radar *(letters/#s, planes)*				radar *(blocks & sessions)*		
Numerical Analysis						radar *(blocks & sessions)*		
Information processing		radar *(varied/consistent mapping, memory load)*, navigation *(message length)*, fusion *(distribution)*	data entry *(fixed v. massed)*	logic decision *(blocked/mixed)*, time estimation *(blocked/mixed)*, navigation *(mixed/blocked length)*		data entry *(variable practice),* radar *(blocks & sessions)*		
Fine motor - discrete		sequence learning *(length, clustering)*	data entry *(fixed v. massed)*			data entry (variable practice)		
Fine motor - continuous								
Gross motor - light								
Gross motor - heavy								
Communication (reading & writing)				foreign language learning (blocked v. mixed), coding (easy 1st v. hard 1st)	foreign language learning (fixed v. expanding items)			
Communication (oral)								

		Practice									
		Task parameters									
IMPRINT task taxons		**Scope (part [e.g. mental rehearsal], default = whole, supplemental)**	**Deep processing (default = no)**	**Mediation (e.g., through prior knowledge)? (default = no)**	**Attentional focus (default = no focus; internal, external)**	**Attentional breadth (default = intermediate; global, local)**	**Stimulus-response compatibility (default = yes)**	**Mapping type (default = consistent; variable)**	**Contralateral training (default = no)**	**Stressor (default = no)**	**Time pressure (default = no)**
Visual			letter detection *(standard/idiosyncratic mappings)*					radar	symbol copy; dart throwing		
Numerical Analysis											
Information processing		data entry *(whole v. partial v. supplemental)*	data entry *(number/words)*, letter detection *(standard/idiosyncratic mappings)*, color naming *(word, sentence)*	fact learning *(person association)*, data entry *(person association)*			data entry *(i/o format); Proctor s-r compatibility*	radar		data entry (hand weights)	memory components
Fine motor - discrete		data entry (whole v. partial v. supplemental)					data entry *(i/o format)*, Proctor s-r compatibility		symbol copy; dart throwing	data entry (hand weights)	
Fine motor - continuous		target finding (mouse reversals)					target finding *(reversals)*				
Gross motor - light											
Gross motor - heavy											
Communication (reading & writing)			color naming *(word, sentence)*								
Communication (oral)											

		Practice			
		Feedback parameters		**Context parameters**	
IMPRINT task taxons		**Presence of (response) feedback (default = no)**	**Feedback scheduling (relative to items)**	**Distractor (default = no; simultaneious , sequential)**	**secondary activity (default = no; simultaneous, sequential)**
Visual				navigation *(noise?)*	
Numerical Analysis				radar *(visual detection)*	radar *(fire decision)*
Information processing		data entry, navigation *(correct/incorrect ; immediate v. delayed)*		time estimation, reconstruction of order, radar *(tone counting)*	time estimation *(letter counting)*, data entry *(articulatory suppression, +/- termination)*, radar *(fire decision)*
Fine motor - discrete		data entry		sequence learning *(tones)*	data entry *(articulatory suppression)*
Fine motor - continuous		target finding *(no reversals; periodic v. trial-by-trial)*			
Gross motor - light					
Gross motor - heavy					
Communication (reading & writing)					
Communication (oral)		navigation	navigation *(abbreviated responses)*		